CALIFORNIA INSTITUTE OF TECHNOLOGY

EARTHQUAKE ENGINEERING RESEARCH LABORATORY

CHARACTERIZING DEFORMATION OF BUILDINGS FROM VIDEOS

BY

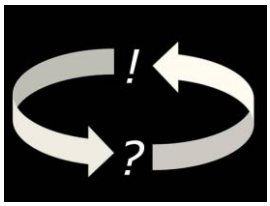
SHERVIN TAGHAVI LARIGANI AND THOMAS H. HEATON

REPORT NO. EERL 2016-01

PASADENA, CALIFORNIA

JULY 2016

This is a collaborative work with STL-Scientific.



http://www.stl-scientific.com

Characterizing Deformation of Buildings from Videos

Shervin Taghavi Larigani[1,2], Thomas H. Heaton[1]

[1] Department of Mechanical and Civil Engineering,
Earthquake Engineering Laboratory,
California Institute of Technology

[2]STL-Scientific

Email: shervin.taghavi@stl-scientific.com

1 Abstract

We have started to explore the feasibility of extracting useful data on the deformation of buildings and structures based on optical videos, (Taghavi Larigani & Heaton).

In the beginning, we look at the characterizations and limitations of the hardware, which is composed of a high-quality digital camera, combined with its optical imaging system capturing a video-footage of the structure under test, and then introduce a straightforward targets-tracking algorithm that produces the time-series displacements of targets that we select on the video.
We performed preliminary measurements consisting of testing our targets-tracking algorithm using high definition format videos displaying the structures that we wanted to test. The measurements pertain to a 1) finite-element software-generated video of JPL/NASA principal building, 2) YouTube-video of a seismic dynamic test of a building, 3) YouTube-video of the Millennium London Bridge "Wobbly Bridge", 4) YouTube-video of a United Boeing 777, 4) YouTube-video of NASA space shuttle rockets during launch.

So far, our tests are encouraging. If our approach proves viable, it can be transformative for the field of earthquake engineering and structural health monitoring. Hence, we consider the prospect of using our technique for surveying buildings and other civil structures in high seismic risk urban agglomerations.

In parallel, the same technique could be applied for 1) real-time structural health monitoring of civil structures, 2) nuclear plants, 3) oil and gas infrastructures, 4) rail & road networks, 5) aircraft, 6) spacecraft, 7) etc., by simply analyzing the structure-facing camera recorded data.

Table of Contents

2 Introduction

We have started to explore the feasibility of extracting useful data on the deformation of buildings and structures based on optical videos displaying the structure under test, (Taghavi Larigani & Heaton). The data that we are looking to extract characterizes the temporal evolution of a structure's spatial deformation. This could lead to the measurement of the displacement field of a building or any other structure as a function of time, which results in understanding the continuum deformation of a building (or another type of structure) as it vibrates. This type of data is the basis of dynamic analyses of a building's mechanical properties, which in the case of large deformations, may vary with time (especially if the structure is being damaged).

Currently, inertial accelerographs are used to infer the deformations of buildings. Instrumenting a building with accelerometers is expensive and requires high maintenance, which means that observations are severely spatially aliased. Each of these sensors typically costs more than several thousand of US$. In order to limit costs, there have been recent attempts to use inexpensive micro-electro-mechanical system (or MEMS) based accelerometers to record motions of relatively dense building arrays (Clayton, et al., 2011). Nevertheless, the complexity of data collection and maintenance of such large net of sensors is substantial. In addition, most building owners will not allow instruments to be installed because of concern about liability, hence only a small number of buildings have been instrumented. In contrast, the operation of several high-performance video cameras should make it possible to characterize the full three-dimensional motions of many buildings and structures in a city. That is, we seek to comprehend the building deformation to deduce how architectural straight lines deform during an earthquake.

Our technique consists of selecting regions of interest of a structure that one would have otherwise instrumented with a series of inertial accelerometers, except that we produce displacement measurements by tracking the motions of recognizable targets with an accuracy that is limited by the camera's physical diffraction limit. Although traditional accelerometer data is typically adequate to measure the vibrations of a structure about equilibrium (approximately linear), acceleration is not well suited to characterize the permanent deformation of a structure

that is structurally damaged. That is, integration of accelerometer data to displacement is difficult, especially if there are tilts of the accelerometer, in which case horizontal accelerometers are sensitive to gravitational acceleration. Thus, displacement measurement is advantageous over acceleration for long integration time, typically starting somewhere from several seconds or tens of seconds, while acceleration measurement is powerful for short integration time, typically smaller than a second. Our overall strategy is to quantify the temporal evolution of the displacement of targets; these displacements can then be used as the input data of structural analysis.

In our approach, there are two issues that can be analyzed separately; one is the characteristics and limitations of the hardware, which is composed of a high-definition digital camera and a lens imaging system, and the other one is the characteristics of a targets-tracking algorithm that we developed to generate the displacements of targets, which are located on the structures we test, as a function of time.

We seek an approach that is robust, in the sense that it is appropriate for diverse characteristics of optical targets. We intend our approach to be as simple, generic, and reliable as possible. We intend to perform the measurement even in difficult environmental circumstances such as irregular changes of background light and luminosity. We also don't assume to have access to the structure that we monitor.

Our preliminary measurement pertains to 1) JPL/NASA's principal building, 2) a seismic dynamic test of a building, 3) the Millennium London Bridge "Wobbly Bridge", 4) a United Boeing 777, and 5) NASA space shuttle rockets during launch. So far, our measurements are encouraging.

If this approach proves viable, it could be transformative for the field of earthquake engineering, and structural health monitoring where the same technique could be used for remote real-time analysis and health monitoring of:

- buildings,
- bridges,
- airplanes,

- spacecraft,
- etc.,

by simply analyzing the data recorded by cameras that face the structure of interest. Additional implementations of that technique include an imaging system installed on the ground and be looking at flying objects when they are in its field of view.

One of our goals consists of deploying in the field autonomous and reliable stations that remotely and passively monitor accurately the mechanical dynamics of objects and transmit processed data to the center. The wealth of information that they could provide is equivalent to extract measurement data from a large series of tens (or larger) accelerometers/seismometers installed in each structure.

In the future, it would be good to explore the feasibility of using optical imagery to independently infer the vibrations of buildings, which have been heavily instrumented. Indeed, in order to understand ground and buildings vibrations, hundreds of accelerometers have been installed on 1) a 52-storey building located in Los Angeles, 2) the NASA's Jet Propulsion Laboratory campus as shown in Figure 1, and 3) the Caltech's campus shown in Figure 2.

We are also contemplating the possibility of utilizing our method for surveying buildings and other civil structures in high seismic risk urban agglomerations. Cities that could most likely be hit by large earthquakes include:

- Tokyo, Osaka, Sendai;
- Taipei;
- Jakarta;
- Manila;
- San Francisco, Los Angeles, Portland, Vancouver;
- Tehran;
- Istanbul;
- Mexico City, Acapulco;
- etc.

Figure 1: Jet Propulsion Laboratory, building 180 from (50 years of talking to space a Nasa social for the deep space network) .

Figure 2: Caltech's Millikan building from (California Institute of Technology). In addition to being heavily instrumented, we have the capability of vibrating the building for experimental purposes.

3 Hardware and Measurement Characteristics

3.1 The hardware

The central part of our detector is a camera. We begin our analysis by considering an ideal case where we model the camera's lens imaging system with a virtual thin lens, of focusing length, f, located at a distance, L_i, from the image plane. We assume that the two-dimensional sensor of the camera has a width designated by D_x, a height designated by D_y , as depicted in Figure 3. By setting the lens for infinity focus, the horizontal and vertical angles of view, α_x and α_y, which are depicted in Figure 4, can be expressed as

$$\alpha_x = 2\arctan\frac{D_x}{2f}, \alpha_y = 2\arctan\frac{D_y}{2f}.$$

Figure 5 displays the fields of view for an ultra-wide angle lens (f=8 mm), a standard lens (f=50 mm), and a telephoto lens (f=300 mm) for a 35-millimeter full-frame digital camera (36 millimeters wide and 24 millimeters tall). These days, with digital cameras using different and generally smaller sized sensors camera manufacturers often quote their lenses or cameras as

having a 35mm equivalent focal length. This is an indication of what the angle of view of a lens would be if it were scaled up to work on a 35mm film camera. As we will see later in an example, low cost entry grade digital cameras have effective focusing length as large as 1200mm.

In the real world, the lens has a finite width that is characterized by its effective aperture A_{eff}, which combined with f determine the maximum cone of light that can enter or exit the lens, as shown in Figure 6. In most cases, the imaging lens is composed of a series of optical elements modeled as a single lens. If all parts of the imaging-lens are considered to be perfect, which means that the optics are enough good quality that flaws in the optical elements (optical aberrations) are negligible, the resolution of the imaging system will then be determined by the wave nature of light, i.e. optical diffraction.

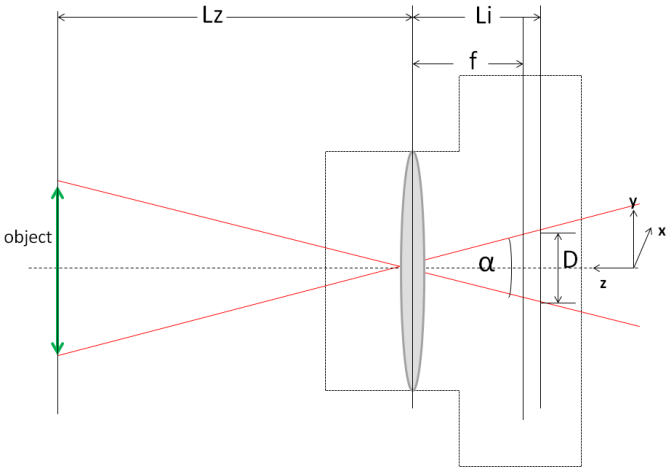


Figure 3: Visual illustration of a digital photography camera from (Wikipedia).

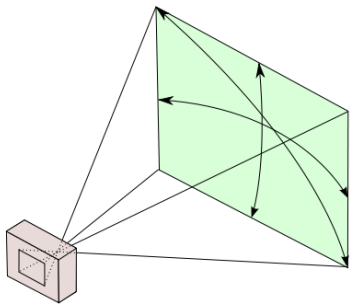


Figure 4: Horizontal and vertical camera's angles and fields of view from (Wikipedia).

.

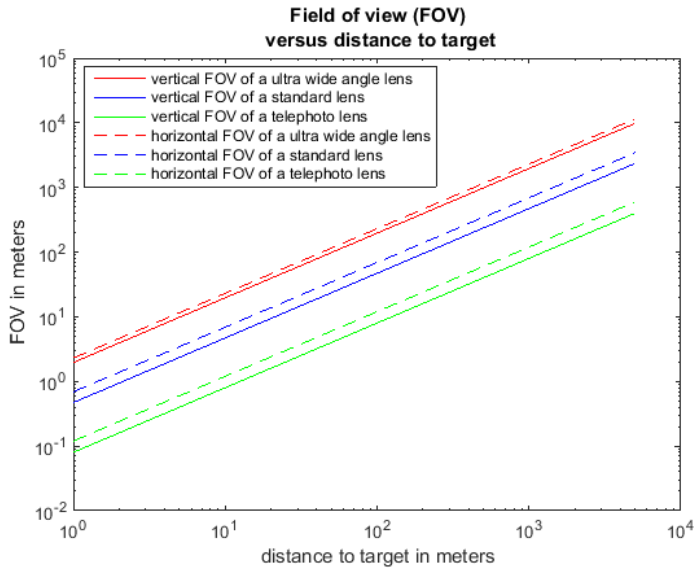


Figure 5: Fields of view of three types of lenses.

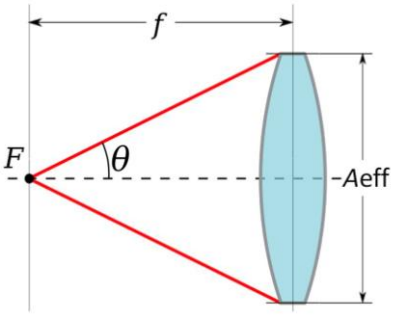


Figure 6: half-angle of the maximum cone of light that can enter or exit the lens

The optical diffraction limit is determined by 1) the finite effective aperture size, A_{eff}, of the imaging lens system, 2) the wavelength, λ, of light and 3) the effective focusing length, f, of the imaging lens apparatus. The diffraction resolution of an imaging system is often expressed in term of angular resolution θ_{min}, which describes its ability to distinguish small details of an object. It is expressed as,

$$\theta_{min} = 1.22\ \lambda\ \frac{N_f}{f},$$

where N_f, the f-number (sometimes called focal ratio, f-ratio, f-stop, or relative aperture) of the imaging lens system, is the ratio of the lens's focal length to the diameter of the entrance pupil, and is given by,

$$N_f = \frac{f}{A_{eff}}.$$

The chart of Figure 7 determines the angular resolution associated with an imaging system. The multiplication of the angular resolution of an imaging system with the distance, L_z, separating the target from the camera leads to the measurement's displacement resolution, , ΔL.

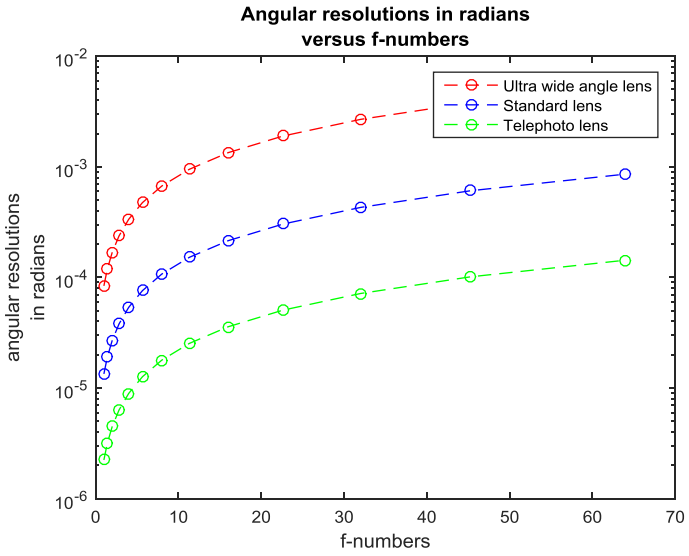


Figure 7: Angular resolutions in radians versus f-numbers for diffraction-limited imaging lens systems.

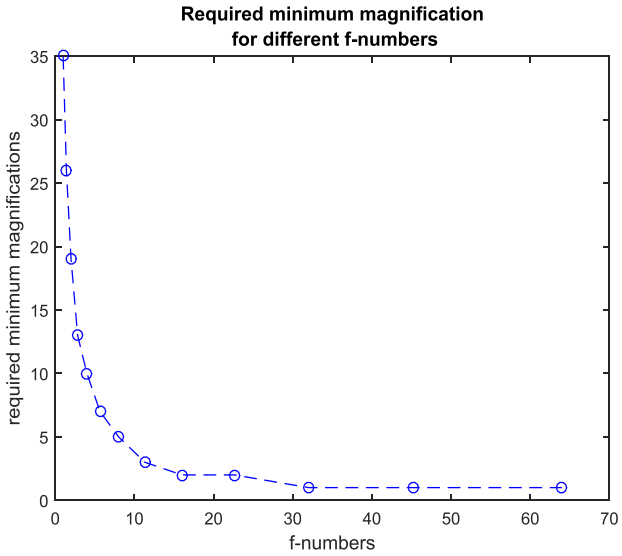


Figure 8: Required minimum magnification for different f-numbers, assuming a pixel size of 6.5x6.5 um.

By looking at Figure 7, we observe that at a distance of several hundred meters from the target we could be able to measure sub-mm displacements using several hundred mm lens imaging system with an f-number smaller than 10. That, of course, requires a robust camera-stabilization procedure (discussed later) and a sufficient depth of field, which is the range of distance in a scene where objects appear acceptably sharp in the image.

Due to diffraction, the response of an imaging system to a point-source is not a point-source but a pattern, whose central figure is the result of the spreading of a point in space, hence the name of Point Spread Function (PSF). The characteristic of the smallest addressable element in an image is determined by the pixel size. Very often the physical size of a pixel on the camera is larger than the size of the image of a point-source on the digital camera's frame. For the camera to be diffraction-limited, it is necessary to enlarge the size of the image on the camera frame of a point source (in the far-field) such it becomes at least larger than two pixels size.

Figure 8 is a chart indicating, for each f-number, the minimum number that the image has to be magnified by; that is a necessary condition for the imaging system being diffracted-limited (for more details refer to Appendix). We observe from Figure 8 that depending on the choice of the

f-number, the minimum magnification factor is between 1 to 35, which is achievable with commercial grade imaging systems.

One of the important characteristics of any measurement in physics is the strain of the measurement; it is also called the measurement dynamic range. The strain of our measurement is defined as the ratio of the smallest measurable displacement in the image, ΔL, over the distance L of the largest object visible in the image, i.e. $\frac{\Delta L}{L}$. It is determined by the number of pixels in each video image. Most digital cameras create high definition video format, which consists of 1920 by 1080 pixels, regardless of their large pixel resolutions. That figure corresponds to a measurement of six parts per ten thousand, or a dynamic range of four orders of magnitude.

In regards to the memory storage, most cameras use a version of secure digital (SD) non-volatile memory cards, either plain old SD, SDHC or SDXC. SDHC stands for High Capacity—meaning between 2 and 32 gigabytes,(Donner, 2014). The secure digital extended capacity (SDXC) allows an even higher memory storage that can go up to 2 terabytes,(Panasonic). Thus, in the ideal case, we could operate a digital camera in its burst mode; that is, continuously and without interruption over a period of few hours (granted that the speed of writing data into the card is fast enough as discussed next).

Parallel to the memory size consideration, the camera's buffer memory size and the speed of writing data from it into the card should be taken into account as well. We have to make sure that the rate of evacuating data from the camera's buffer is faster than the rate of upcoming data into it in order to avoid any sort of congestion.

By all means, memory storage and management are critical, any upcoming system design will have to take them into account

The choice of the shutter speed (distinct from frames rate), which is is the length of time a camera shutter is open to let the camera sensor being exposed to light, is important. A slow shutter speed causes blurring of the image along the direction of motion of a moving object. On the other hand, if the shutter speed is fast it can help to freeze the action; thus, the image will appear sharp at the expense of less light-illumination.

3.2 Error analysis

In this section, we look at the different parameters of the imaging system and examine their impacts on our measurement.

The camera is not an inertial frame; we have to minimize the impact of its motion on our measurement as much as possible. That is why camera anti-shake technologies are often integrated into the imaging system; we talked about them in the Appendix. Alternatively, professional photographs minimize the outcome of shaking on the image by reducing the shutter speed (without using anti-shakes). The empirical procedure they use indicates that the shutter speed should be much smaller than the lens focal length, (Browne). For example, the shutter speed would be 1/100th sec or faster when using a 50mm lens.

However, as the shutter speed increases the time for which the sensor is exposed to light decreases; therefore, less light is available to take images. To go beyond that limitation, the level of sensitivity of the camera to available light, ISO, can be increased, which corresponds to the amplification of the electric signal produced by the sensor. The imaging sensor's sensitivity is proportional to the ISO number. The error that the electronic noises induce in the images is negligible as long as the ISO number is small, typically lower than 400 or 800.
ISO numbers start from what is called the "base ISO", which is the lowest ISO number of the sensor that can produce the highest image quality without adding noise to the picture; the series of ISO numbers are 110, 200, 400, 800, 1600, etc.
High ISO comes at the price of adding noise to the images, which becomes visible, typically, for ISO larger than 800.
The noises in an image sensor are:

- Shot noise

  It is the photon counting noise of the optical sensor; it is caused by the particle nature of light.

- Read noise

  It is a combination of all the noise generated by the conversion of the electric charge of each CCD pixel into an electric signal.

- Dark noise

  It is the electron shot noise of the sensor; it consists of the statistical variation in the number of electrons thermally generated within the pixel independently of the light that illuminates the pixel sensor.

However, we think that our algorithm reduces (or even cancels) the impact of the ISO noise as long as the targets are plainly visible in the images.

Understanding the consequences of image compression algorithms on our analysis is important (Donohoe); they remove image data that is recognized as redundant or irrelevant. Nevertheless, the same data might be relevant to our technique. That is, we would like to work as much as possible with uncompressed video images that contain information for every pixel, in every frame. Nonetheless, the file size of such data is large (that is why the raw data is compressed). Consider the raw data of a black-and-white high-definition (HD) format video, consisting of a frame of 1920 by 1080 pixels. Each pixel typically consists of 8 bits (1 byte). That is, the raw data file size of a single black-and-white image of 1920 by 1080 pixels is 2.07 megabytes.
High-quality commercial digital cameras are capable of producing uncompressed videos in the HD format at a rate of more than 11 images per second. Alternatively, the camera can operate in a photography burst mode where photographs are captured in quick succession (often at a rate of 11 images, or more, per second). That corresponds to a data rate of 220 megabytes per second for a camera resolution of 20 megapixels capturing grayscale images.

Image distortion is another mechanism that can cause an error in our measurement. In most cases, the image captured by a camera is distorted. The distortion can be created by,

- the inherent fabrication imperfections in the elements that constitute the lens imaging system, that is named optical aberrations;
- the perspective distortion of an image, which is related to the focusing length of the imaging system.

Effects of optical aberrations on the images can be removed by calibrating the imaging system with a reference image where all the characteristics of the features that are displayed in it are known. Such algorithms exist; an example is mentioned in the Appendix.

On the other hand, perspective distortions are independent of the quality of the optical imaging system; they represent the distortion of the representation of objects and of their relative sizes when they are displayed in an image. They are categorized as either extension or compression distortions.

Typical of telephoto lenses (large focusing lengths), compression distortion is observed in images where objects that are more distant from the lens emerge as the same size than the objects that are closer to the lens. Hence, a person who looks at the picture does not distinguish the distance separating the objects along the range; that is, the image seems to be compressed along the range of the camera.

Typical of wide angle lenses (small focusing lengths), extension distortion is observed in images where objects closer to the lens appear much larger than the objects that are more distant from the lens. Hence, a person who looks at the picture perceives that the objects along the range of the camera are more distant from each one than what they are in reality.

In addition, light rays, which should be straight, begin to curve at the edges of an image; this phenomenon is particularly noticeable in the case of short focusing lengths that characterize wide-angle lenses and is known as the fisheye distortion.

Perspective distortion induces changes in the scale in the images that we use for measuring motion; hence, we need to know the dimensions of the structures that are represented in the images in order to correct for the error that perspective distortion induces on our measurement. However, dimensional characteristics of most structures and objects are easily accessible nowadays. Even in the peculiar case of buildings, we can figure out their dimensions using Google Earth. As a consequence, we don’t necessarily need to have access to the structures.

Considering that we are interested in measuring the motion of one target relative to the motion of another target (i.e. a differential measurement), errors that are common to both measurements cancel out. Typical of optical and vision distortions, let's assume that the location of the image of a target relative to the image of other objects in the camera frame is different from reality; yet, as long as the imaging system tracks flawlessly the displacement of the target, there is no error that

is induced in our measurement. This type of distortions could occur naturally due to environmental characteristic changes of the medium of propagation of light rays coming from the targets.

In the case of buildings, we are seeking to measure displacements with a precision of a cm or less. Consequently, the index of refraction of the medium of propagation of light rays should change by a significant amount over a range of m in the direction of motion of a target for the medium to induce such error.

As an example, the DSC-HX300 (Sony) entry-level digital camera has a varying effective focusing length that can be as large as 1200 mm; we estimate that a unit pixel represents a length of one cm at the field of view when the targets are at a distance of 1.84 km from the camera as explained in more detail in the Appendix.

The optical path is the path that light travels from the target to the camera; if it changes during the period of the measurement, an error is induced in the measurement of the displacement of a single point. A change in the optical path can be caused by a change in the index of refraction, which characterizes how light propagate in the medium. Indeed, the optical path length is the product of the geometric length of the light path and the index of refraction of the medium in which light propagate. The index of refraction is a function of many variables, which include the temperature, the pressure, and the density of charged particles in the medium. So, changes of these variables of the air cause changes in the index of refraction, which, in turn, induces changes in the optical path.

However, in the case of a differential measurement, as in our case, the error cancels when it is common to both measurements. So, changes in the optical path do not affect our measurements as long they are the same for all light rays coming from the structure under test.

Environmental conditions often induce irregular changes of the luminosity of the targets as perceived from the camera. But, the straightforward displacement measurement algorithm that we have developed, which is described in the next section, is not sensitive to these changes as long as the targets remain visible.

On the other hand, environmental conditions can lower the signal to noise ratio of the measurement by blurring the images that are recorded by the camera. That is, for example, the case when the light coming from the targets get scattered, e.g. the light rays are deflected from a straight line path by irregularities (e.g. impurities or particles) in the propagation medium.

In a similar way, video and image compression algorithms can deteriorate the measurement signal to noise ratio, by reducing the quality of the images.

Optical mirages are also phenomena that could reduce the signal to noise ratio of our measurement. They are optical phenomena in which gradients in the refractive index of the air create distinct layers of air through which the rays of light travel before reaching the camera. The interaction of hot and cold air, which is called convection, causes variations in the temperature of the air; that, in turn, creates a gradient in the air index of refraction. That is, the paths of the light rays bend. However, the bending of light by itself is not of great concern since all light rays coming from a particular distant object travel, almost, through the same optical path. However, convection, which causes shimmering, is a turbulent mechanism, which is characterized by unpredictable chaotic changes; hence, it is an unstable process that degrades the quality of the image by blurring it.

So, our technique works as long as we can recognize the targets in the image. So as to diminish undesirable consequences of shimmering on our measurement, we can select a shutter speed that exceeds the rate at which the index of refraction of air changes.

4 Targets-tracking algorithm

Feature tracking and motion detection algorithms have existed in the area of computer vision. In the future, we are interested in investigating the feasibility and limitations of these algorithms (or modifying them) to detect buildings' motions. However, we began to develop a straightforward targets-tracking technique that seems especially suitable for the sort of applications we are interested in. As explained in the introduction, we intend our approach to be as simple, generic, and reliable as possible. We seek an approach that is robust, in the sense that it is appropriate for diverse characteristics of optical targets. We intend to perform measurements even in difficult environmental circumstances such as irregular changes of background light and luminosity of the

targets as perceived from the camera. In addition, we don't assume to have access to the structure that we monitor.

4.1 Our approach

Analogous to image template matching technique, we have chosen to utilize the basic normalized two-dimensional cross-correlation function to determine the position of the (2x2) matrix $\Im_2(x,y,t_k)$ inside the (2x2) matrix $\Im_1(x,y,t_n)$. $\Im_2(x,y,t_k)$ is the two-dimensional matrix representation of the digital image associated with the image of a target extracted from the frame number k in the video images sequence. Similarly $\Im_1(x,y,t_n)$ is the two-dimensional matrix representation of the digital image associated with the video frame number n in the video images sequence. By finding the pair of η_x and η_y coordinates that maximized,

$$\gamma(\eta_x,\eta_y,t_n,t_k)=\frac{1}{N\sigma_1\sigma_2}\sum_{x,y}\left(\Im_1(x,y,t_n)-\overline{\Im_1(t_n)}\right)\left(\Im_2(x-\eta_x,y-\eta_y,t_k)-\overline{\Im_2(t_k)}\right), \qquad 1$$

we calculate the location of the target inside the image frame number k, where *N* is the number of pixels in $\Im_2$, σ_1 is the standard deviation of $\Im_1$, $\overline{\Im_1}$ is the average of $\Im_1$, and similarly σ_2, and $\overline{\Im_2}$ represent the same quantities relative to $\Im_2$. The search domain can be reduced to a sub-region of the video-image as long as the target stays inside it during the course of the video. This approach has the advantage of being invariant to global brightness change in $\Im_1$ and $\Im_2$ caused by exposure and luminosity variation conditions during the course of the film footage since the mean image brightness is subtracted from each pixel value.

In our applications the camera settings do not change over the course of the movie, and since the targets are located in the camera's far-field, changes in the image scaling and perspective distortions are therefore generally minor during the period of the movie. One difficulty with our

approach occurs if $\mathcal{G}_1(x,y,t_n)-\mathcal{G}_1(x,y,t_k)$ cannot be neglected. In particular, there is the possibility that diverse pairs of (η_x,η_y) that have no physical correspondence to each other may still maximize $\gamma(\eta_x,\eta_y,t_n,t_k)$. That is, identification of features may not be unique. To mitigate this problem we constrain our selection in the pair of coordinates such that it 1) maximizes γ and, 2) correlated points should appear reasonable to a human checking the analysis. We also observe that by properly selecting the targets, we were able to considerably reduce unwanted artifacts that sometimes arise using our cross-correlation methodology.

The two-dimensional cross-correlation is rotation-variant, which means that $\gamma(\eta_x,\eta_y,t_n,t_k)$ changes as $\mathcal{G}_1(x,y,t_n)$ rotates relative to $\mathcal{G}_2(x,y,t_k)$. Nevertheless, this characteristic is of great interest to us; it allows us to detect spatial rotation of a target. We expect to discuss the issues associated with target rotation in a future publication.

One method of automatically reducing the number of nonphysical solutions (reducing the need for human intervention) and for also prevailing over the rotation no-invariance characteristic of cross-correlation consists of updating $\mathcal{G}_2$ at the end of each time iteration n, so

$$\mathcal{G}_2(t_n)=\left(\left(\eta_x,\eta_y\right)_{\max(\gamma(t_n,t_k))},\mathcal{G}_1(t_n)\right),$$

and at the next iteration n+1 uses the most recent version of the image of the target, $\mathcal{G}_2(t_n)$, to maximize

$$\gamma(\eta_x,\eta_y,t_{n+1},t_n)=\frac{1}{N\sigma_1\sigma_2}\sum_{x,y}\left(\mathcal{G}_1(x,y,t_{n+1})-\overline{\mathcal{G}_1(t_{n+1})}\right)\left(\mathcal{G}_2(x-\eta_x,y-\eta_y,t_n)-\overline{\mathcal{G}_2(t_n)}\right) \quad 2$$

For future works, we intend to analyze the evolution of $\mathcal{G}_2(t)$ over time, which would form the basis of a Eulerian analysis of the target complementing our current Lagrangian approach consisting of monitoring the dynamics of the targets relative to each other.

4.1 Existing motion and tracking algorithms

Normalized cross-correlation is ordinarily not the ideal approach to feature tracking since it is not invariant with respect to image scaling and perspective distortions. An entire field of computer vision is dedicated to the non-trivial problem of recognizing and matching alike features displayed in different images. One of the challenges that many have been working on consists of recognizing the same feature in two different images and succeeding in matching the different views of the same object. Such applications require the algorithm being scale-invariant, meaning it is capable of recognizing the same feature at different scale length representations. A well-known algorithm is the scale-invariant feature transform (or SIFT), (Lowe, 1999). It consists of first establishing an "image descriptor" of an object by extracting its distinctive features from a set of images depicting the object. The "image descriptor" describes the elementary characteristics of the object. The descriptor is then used to identify the object when attempting to locate it in a different image containing many other objects. Since it is important that the extracted features being detectable even under changes in image scale or rotation, the process of object-matching consists of calculating the relative location of a set of features of one object and compare to another.

More recently, a variant called speeded up robust features (or SURF) (Bay, Tuytelaars, & Van Gool, 2006) exhibiting faster features extraction appeared to improve SIFT slow processing time. Another similar descriptor is the histogram of oriented gradients (HOG) (Dalal & Triggs, 2005), which counts occurrences of gradient orientation in localized portions of an image. (Histogram of oriented gradients). To our knowledge SIFT, gradient location and orientation histogram (or GLOH) and their variants exhibit the highest matching accuracies compared to other images descriptor algorithms. To comprehend the limitations and challenges of such complex algorithms, one could consider that a reliable image search engine similar than current word search engines such as Google or Yahoo has not yet been realized; that is still an ongoing problem. Moreover, such algorithms have not been designed to provide the range of accuracy and precision we are looking to measure, hence they might not be best suited for our application.

These algorithms increase complexity; each one featuring distinctive attractive and unsatisfactory attributes. However, in future works, we would like to investigate how variations of existing image descriptors could be applied to our problem.

Currently, image-processing techniques commonly designated as super-resolution can achieve a detectable displacement motion smaller than the limit imposed by the optical characteristics of the imaging system, i.e. the diffraction limit, (Betzig & Trautman)(Born & Wolf)(Cox & Sheppard, 1986)(Fox, 2007)(Gustaffsson, 2000)(Lukosz, 1966)(Westheimer, 2012)(Zalevsky & Mendlovic) (Harris, 1964)(Di Francia, 1955)(Poot, et al., 2012)(Simpkins & Stevenson, 2012)(Farsjiu, Robinson, Elad, & Milanfar, 2004)(Schuon, Theobalt, Davis, & Thrun, 2009). Nevertheless, they can only operate under specific luminosity and background light conditions that are not always consistent with the ambient environment of videos we wish to analyze (e.g., building deformation in an earthquake).

A group at MIT has already been investigating similar problems as algorithms utilized for super-resolution and they developed "the Eulerian video magnification technique", (Wu, et al., 2012). Their approach consists of amplifying subtle light-intensity variation in the video frames time series happening at a temporal frequency selected by the user. By choosing a temporal filter tuned to human heart rates, they were able to visualize the variation of redness as blood flows through the face. Since the signal they look for is smaller than the inherent video noise, they apply the temporal filtering to low spatial frequencies (spatial pooling), which consists of regrouping together local features from spatially adjacent pixels. Typically, this is done to improve sensitivity to small deformations of objects. Hence, their technique is suited for analysis of low image spatial frequencies (smooth images); it is not designed to analyze high spatial frequencies (high-contrast images). They used the same technique to magnify low-amplitude motion such as the subtle motions around the chest of a breathing baby. Nevertheless, their assumptions hold for small motions, essentially for small perturbations. They rely on the same differential approximations (i.e. small motion approximation) that are the underlying foundation of optical flow algorithms, (the concept of optical flow is described in more detail in the Appendix). A potential weakness of this approach is that, in practice, it seems to be sensitive to changes in background luminosity. For example, if a user located in the US uses a high-speed

camera and selects a sampling rate that is a rational fraction of 60 Hz, then the output video may display a color/motion magnification of artificial phenomena caused by the flickering of artificial lights driven by AC electric power.

This research came after an earlier work on the subject of motion magnification produced from the same laboratory in MIT,(Lui, Torralba, Freeman, Durand, & Adelson, 2005). It consists of identifying clusters in an image and then regrouping objects based on their similarity in position, intensity, and motion characteristics. The user chooses which cluster's motion should be amplified and by how much. The analysis assumes that the input image sequence depicts a predominantly static scene on a static background (unlike an earthquake occurrence). Once again, their goal consists of amplifying the visualization of the motion of a specific cluster in the image.

Several other research groups have begun to investigate the use of video camera images to detect the horizontal vibrations of bridges by placing targets on to them. (Zhou, Xiab, Weib, & Wu, 2012) placed solid black circle targets on the white color selected cable of the China Jing Jiang Bridge, while (Chong, Al-Baghdadi, & Alshadli, 2014) placed retro-reflective targets beneath a bridge, illuminated them with near-infrared illuminators, and used near-infrared cameras to detect the horizontal motions of the targets. In (Zhou, Xiab, Weib, & Wu, 2012) the targets single themselves out from their vicinity by not reflecting light, while in (Chong, Al-Baghdadi, & Alshadli, 2014) the opposite occurs.
In our approach, we would like to monitor a structure without having a physical access to it.

5 Measurements

In this section, we perform measurements on variety of videos and then conduct the analysis. We demonstrate how the technology might be used in a variety of experiments.

In the case of the YouTube videos, we don't have any information about the characteristics and settings of the cameras that recorded the footages; nor do we know what kind of movie compression schemes have been applied to them.

In all these cases, the video files that we worked with were compressed. As discussed earlier in " 3 Hardware and Measurement Characteristics", video compression programs remove any data (of the video) that is considered as redundant or irrelevant. Nevertheless, the same data might be relevant to our technique. That is, an uncompressed video file is more suitable for our approach and easier to work with.

Regarding the videos that were downloaded from YouTube, they went through two distinct steps of compression in all likelihood. First, the raw data of the video images were compressed to the video file that the camera generates at its output. The same video file was compressed a second time during the download on YouTube. Unfortunately, we don't have information about the compression schemes that they underwent.

Unfortunately, the compression schemes cut by an unknown amount the information relevant to our technique. Hence, if we had access to the raw data of the videos, we would be able to detect displacements at a precision that are not detectable when the videos are extracted from YouTube. Another concern is that video compression schemes might actually induce errors (that we are not aware of) in our measurement.

To reduce the likelihood of such mistakes, we tested our measurements to see whether they physically make sense. Nonetheless, we cannot state that our measurement is free of possible errors induced by the video compression programs. However, we were able to use our analysis on a video simulating the motion of a building using finite element analysis to establish that creation of the video did not introduce artifacts of importance for our analysis.

In addition, all the YouTube videos analyzed in this section possess perspective distortions.

5.1 Simulations of JPL/NASA 180 building

In this section, we compare our method of measurement with a known result in order to test the reliability of our method. That is, we compare measurement performed by applying our technique to a video displaying the structural dynamic of JPL/NASA 180 building (shown in Figure 1) with the data used to generate the video.

Building 180 is currently a study area of the Community Seismic Network (CSN) project. 27 3-component accelerometers have been installed in this nine-story steel building. Structural blueprints of the building were used to construct an ETABS model. Based on that model, (Massari) generated a video simulating the motion of the building using the Extended Three-dimensional Analysis of Building Systems (ETABS) structural and earthquake engineering finite-element software, with a precision of sub-nanometer; Figure 9 depicts a view of the JPL/NASA 180 building generated by ETABS. The data, which generated the simulation, corresponds to what was measured by the CSN at the building base for the Fontana earthquake, which struck Southern California on July 25, 2015, with a magnitude of 4.2.

In turn, we used the video to record the displacement of the nodes and then compared our measurement with the ETABS generated data. In the following analysis, we express all the displacements relative to the base of the building, which is a non-inertial reference.

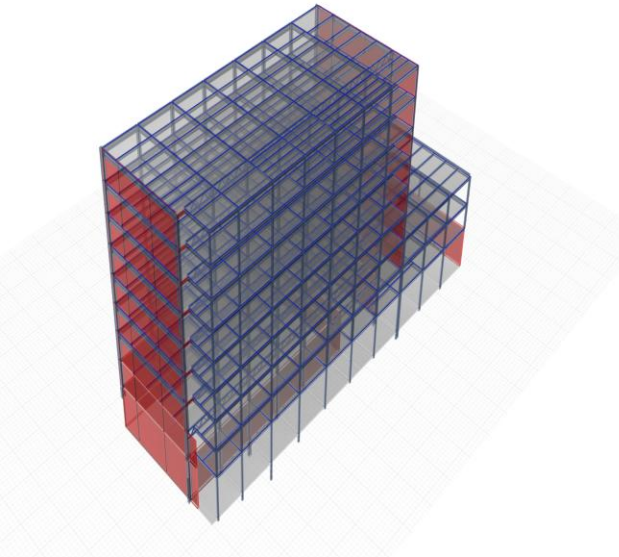


Figure 9: Three-dimensional view of JPL/NASA building 180 generated by employing ETABS, (Massari) .

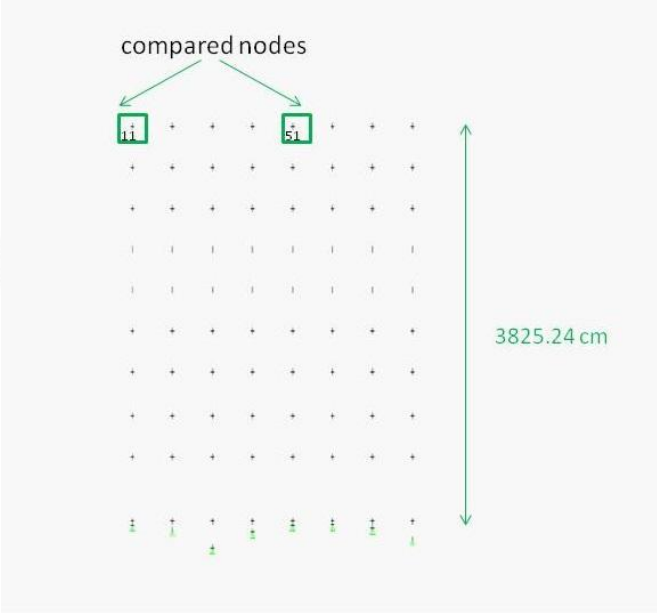


Figure 10: ETABS schematic of the front view of building 180. The two nodes used for comparison are displayed in green.

For the purpose of comparison, we used the front view of the building decomposed in 8 nodes in the horizontal direction by 9 nodes in the vertical direction as displayed in Figure 10.

The video has an image resolution of 1.7 megapixels with a frame rate of 30 images per second; however the data provided to generate the video has a sampling frequency of 50 Hz.

We deduce the dimension that each pixel represents by taking into account the physical height of the building (3825.24 cm) and dividing it by the number of corresponding pixels in the video frame (559 pixels). Henceforth, we estimated that each pixel represents a length of 6.84 cm and dividing it by the magnification factor (x10000) we obtain the correct measure of the displacement.

5.1.1 Data comparison

For this writing, we chose to display the comparison of the dynamic of the two nodes shown in Figure 10. The top plot of Figure 11 represents the generated ETABS displacement of the target (11), whereas the bottom plot describes the measurement of the displacement of the same target

following our technique. Likewise, Figure 12 represents the comparison of the data generated by ETABS located on the top of the figure with our measurement located on the bottom.

We estimated earlier that each pixel represents a length of 6.84 cm and we used a magnification of ten thousand. As a result, 6.84 um is the smallest displacement we can measure; that is four orders of magnitude larger than the resolution of the data generated by ETABS. In spite of possessing a measurement dynamic range (i.e. the difference between the smallest and largest measured displacement) four orders of magnitude smaller than the data generated by ETABS, we have been successful to reconstruct from the video of the building the dynamic behavior of each node confirming with great precision to the initial signal even at the higher frequencies. When we zoom in on any video image, we observe blurriness and imperfection in the reproductions of the characteristic features of the model. That might be an artifact of the video compression format. However, our measurement matches well the known motion of the building; thus, we conclude that the video compression didn't affect our measurement.

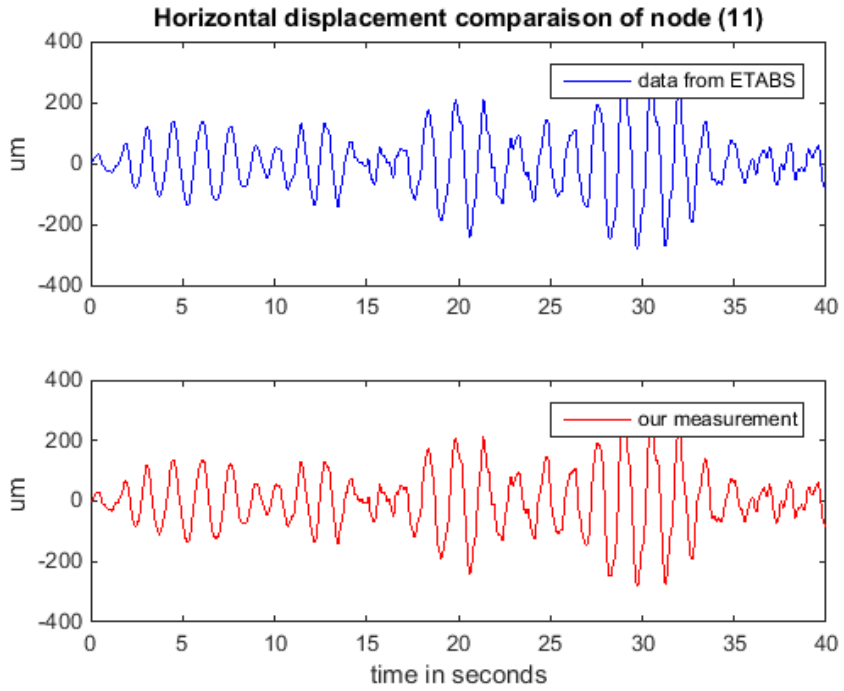


Figure 11: Horizontal displacement comparison of node 11. The top plot represents the data generated by ETABS; the bottom plot represents our measurement. Both data are expressed relative to the motion of the base of the building.

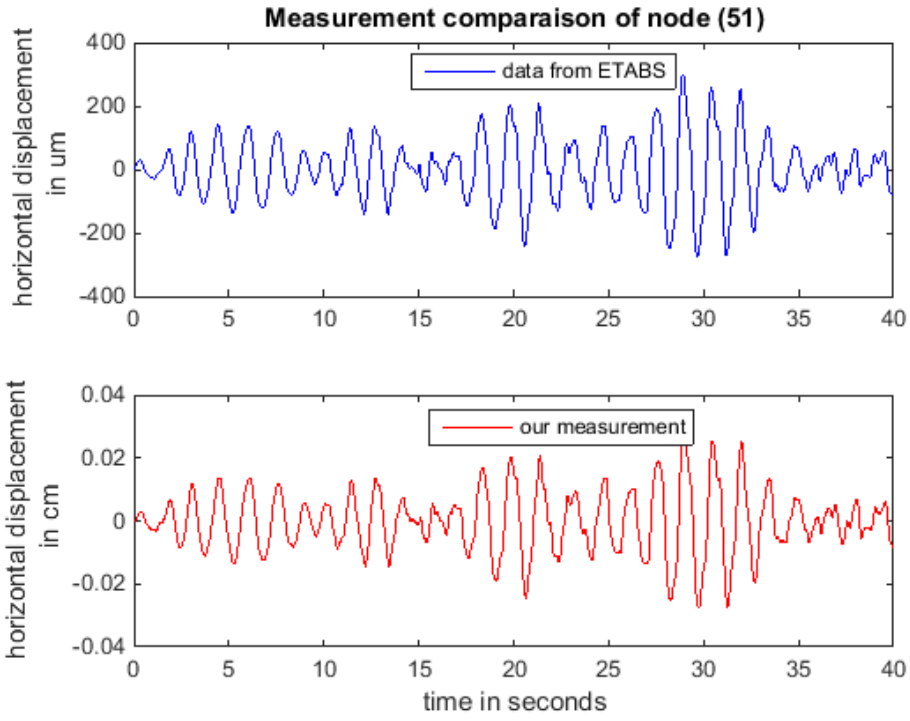


Figure 12: Horizontal displacement comparison of node 51. The top plot represents the data generated by ETABS; the bottom plot represents our measurement. Both data are expressed relative to the motion of the base of the building.

5.1.2 Some application

Analysis of videos could be used to characterize deformations from diverse physical phenomena. From example, the distance measurement between two nodes of a building can be used to estimate the second strain invariant, which is a scalar measure of the shear. The following discussion demonstrates how videos can be used to investigate the modal proprieties of buildings.

Figure 13 represents the horizontal displacement power spectral density of node (11).

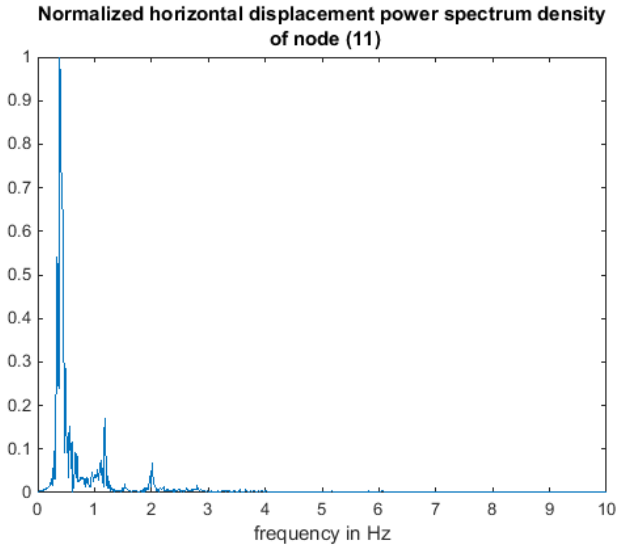


Figure 13: Normalized horizontal displacement power spectrum density of node (11) for the motion shown in Figure 11 .

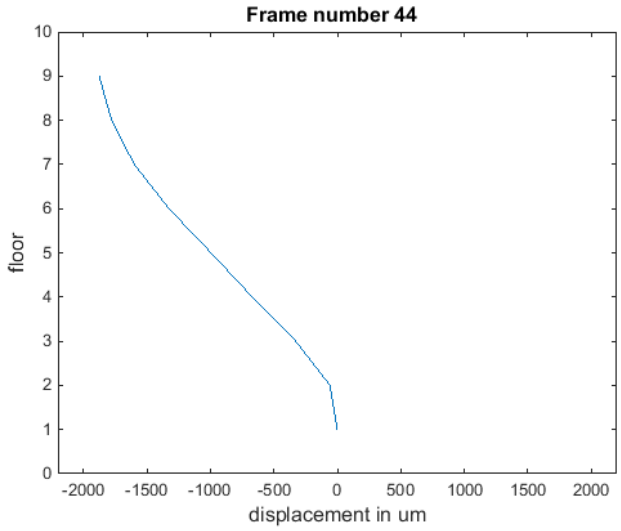


Figure 14: Snapshot of the temporal modal dynamic of the left column when vibrating at a frequency of 0.38 Hz with a decay rate of 4%.

It seems that the building's first mode is around 0.38Hz, the second mode around 1.184 Hz, which is about three times the frequency of the first mode, and the third mode at 2 Hz, which is about five times the frequency of the first mode. This relationship between the first modes is similar to the way in which the modes of a continuum shear-beam building are connected to each other (Cheng & Heaton, 2015). But, real buildings are not a continuum, and their modal shapes are not exactly the same ones of a continuum building. As a side note: moment-resisting-frame

and shear-beam buildings have similar resonant modes. We generated a video displaying the left column's temporal modal behavior. Figure 14 is a snapshot of the video when we applied a temporal filter at a frequency of 0.38 Hz with a decay rate of 4% corresponding to what appears as the main resonant frequency of the building. Figure 15 is a snapshot of the video when the frequency was set to 1.184 Hz, corresponding to what appears as the building's second vibration mode. Figure 16 is the film snapshot when the frequency was set to 2 Hz; the estimated third modal vibration.

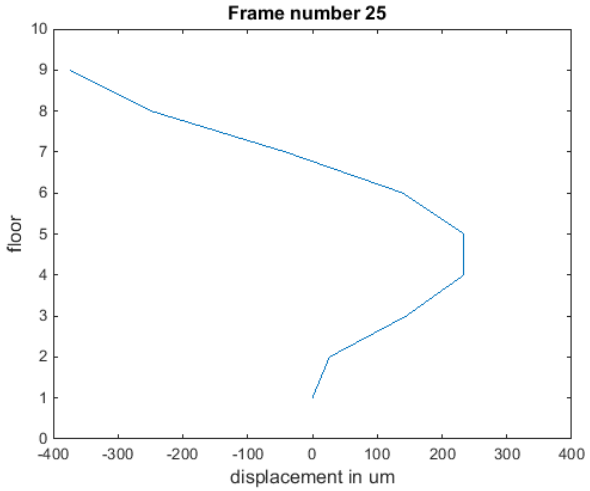


Figure 15: Snapshot of the temporal modal dynamic of the left column when vibrating at a frequency of 1.184 Hz with a decay rate of 4%.

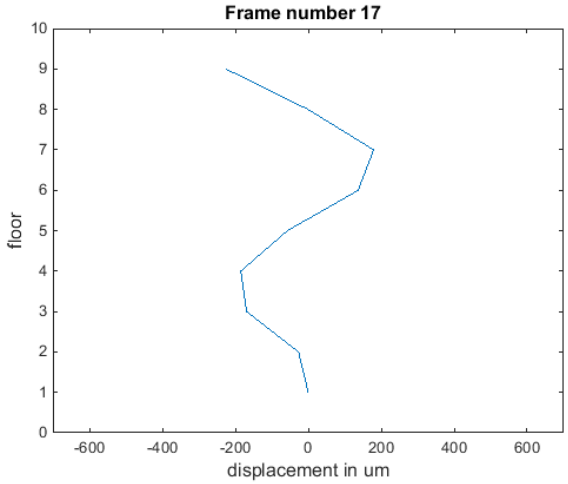


Figure 16: Snapshot of the temporal modal dynamic of the left column when vibrating at a frequency of 2 Hz with a temporal decay rate of 4%.

5.2 Seismic dynamic test of a building

In this section, we are interested in applying our technique of measurement to a video depicting a building downloaded from YouTube, (E-defense Dynamic Test of Seismic Isolated Building by Sannomaru Eq.), as displayed in Figure 17. The targets are shown in white and are labeled as ij ; i is the column number and j is the floor number. The column on the left side is denoted as 1, the next column on the left side is denoted 2, etc. Likewise, the first floor is denoted 1, the second floor 2, and the third floor 3. So, the target on the left side at the third floor is denoted 13.

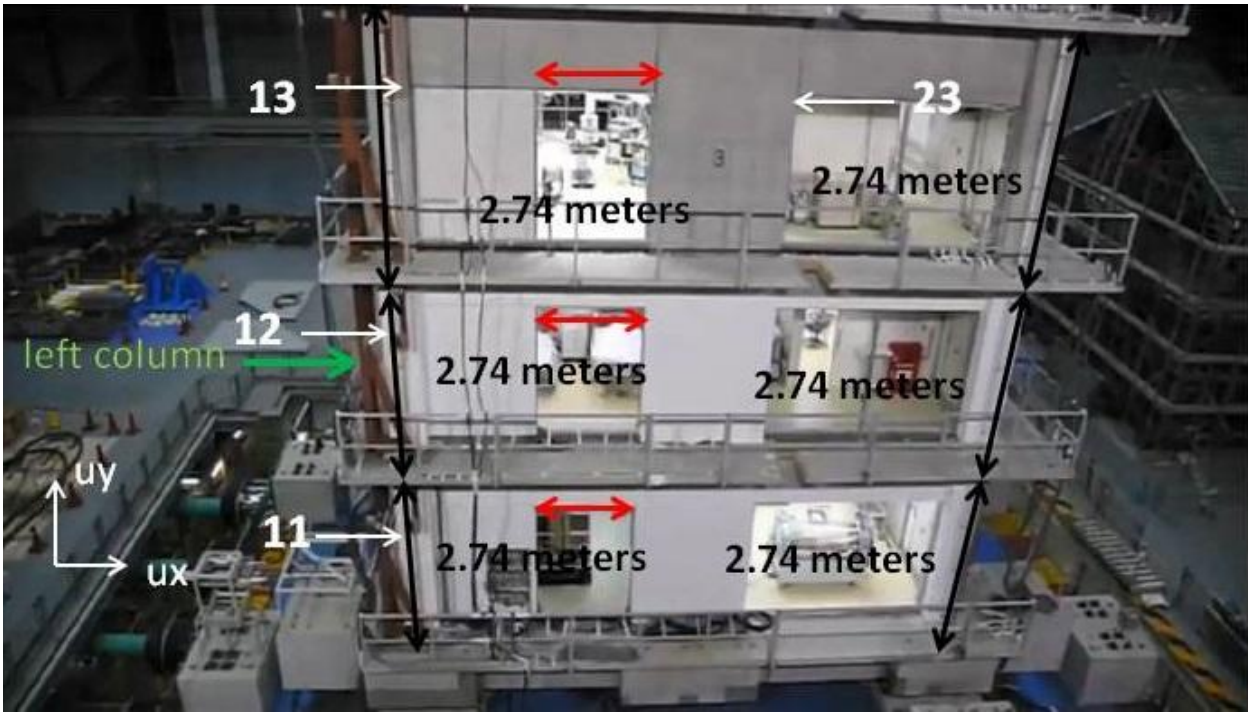


Figure 17: View of the building. The targets are in white. The black bars are the floors ceilings, which we assume being similar. The red bars are the widths of openings, which we assume being similar.

This example exhibits several interesting adversities consisting of:

- perspective distortions,
- sub-regions of the building coming in and out of focus during the course of the video,
- the building experiencing large three-dimensional motions,

- low video images resolution consisting of 0.92 megapixels at a frame rate of 25 images per second.

To test the stability of the camera, we chose references located away from the building that might not have been affected by the motion of the edifice. Nonetheless, their measured locations change over the course of the video footage, thus we cannot assume the camera that took the video was stationary. To prevail over this complication for performing a thorough analysis, one might pick an arbitrary reference that could be located as an example in the lower part of the building, and then study the motions of the rest of the building relative to its lower section.

To estimate the scales on the image, we assume that each floor has the standard ceiling height of 9 feet (2.74 meters as indicated by the black bars); consequently, we deduce the physical dimension of our measurement. Nonetheless, on account of the perspective distortion, the physical pixel representation varies along the horizontal and vertical axis of the camera. We approximate that along the vertical axis of the image along the column located on the left side of the building an image pixel represents 0.77 cm on the third floor, 1.18 cm on the second floor, and 1.2 cm as shown in Figure 17. In like manner we evaluate a pixel along the vertical direction of (21) representing a distance of 0.84 cm.

We suppose that a pixel in the high part of the third floor represents the same distance either in the horizontal or vertical directions of the image; that is, of course, a coarse approximation since the scaling in the image is not isotropic. Thus, we evaluate the width of the open entrance closest to the left side of the building. We also observe that similar to the third floor; similar entrances are located on the second and first floors (indicated by the red bars). If we assume that the three of them are identical to each one, then they should have identical size. Hence, by measuring the width length of each entrance relative to each other, we calculate the ratio by which they have been scaled from one floor to the other; thus, we estimate the distance that a pixel represents in the horizontal direction on each floor. Therefore, we approximate that along the horizontal axis of the image, $\overrightarrow{u_x}$, on the left side of the building an image pixel represents 0.77 cm on third floor, 0.84 cm on second floor, and 0.92 cm on the first floor.

In this writing, we display result pertaining to the dynamic of the targets shown in Figure 17. Figure 18 displays the horizontal displacements of the targets located on the left column of the edifice, and Figure 19 displays their vertical displacements.

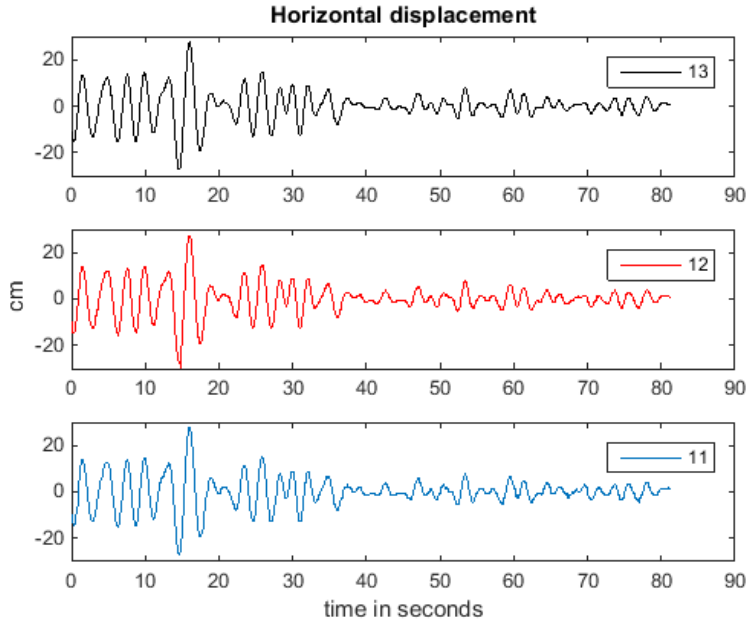


Figure 18: Horizontal displacements of the targets along the left column.

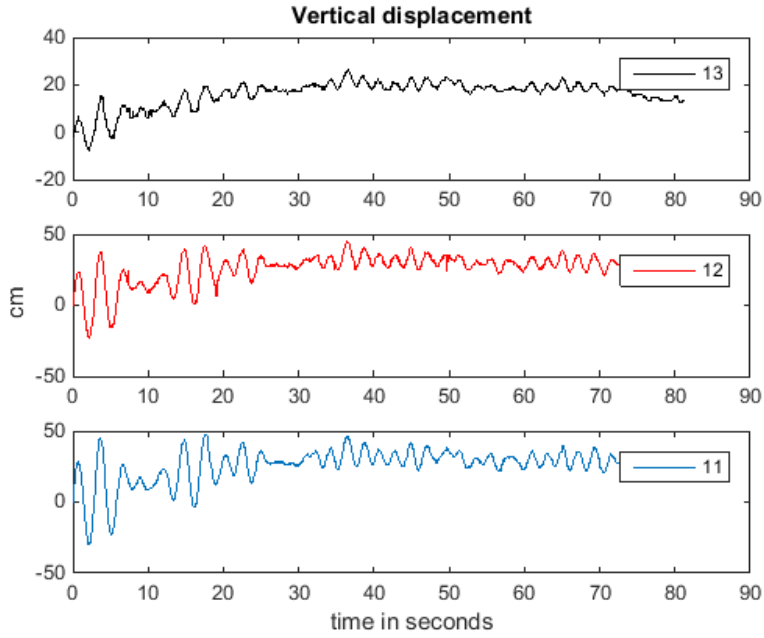


Figure 19: Vertical displacements of the targets along the left column.

Conventionally, we would have to display the measurements relative to the motion of the base of the building, thus subtracting the motion of the base from the measurements that are

displayed. Nonetheless, as discussed earlier, we don't know the correct dimensions of the building nor do we know exactly the orientation and position of the camera relative to the building.

We calculated the scales in the image, approximately, to infer a coarse estimate of the orders of magnitude of the displacements we measured. Nevertheless, there likely exist errors in the variations in scale, which we estimated, from one floor to another one and from a side of the building to another side. That is, in order to avoid inducing fictitious patterns in the time-series data, we don't express our measurements relative to the motion of the first floor of the building.

The data of all the vertical displacement measurements ends in an unusual way as shown in Figure 19, none of them return to its initial position when the time series stop; it seems like that all of them experience similar drift.

- One reason might be related to the fact that the video footage that we have analyzed is an 81-second-long video that we selected off the YouTube video. This selection is justified by our motivation to work with a video where the field of view does not change during the period of the footage. In the section of the video that we singled out, the building is still in motion when the footage stops. That might explain the drift that we observe in the data of the vertical displacements.
- Another explanation of the drift could be related to the imposed motion on the building, which could also have a vertical component into it; thus, we could imagine that the building would have been elevated since the footage started.
- Another explanation of that variation in height might be related to a hypothetical displacement of the camera during the footage since we were not able to confirm whether the camera that took the footage was stationary.

Nevertheless, it seems that the vibrations we measured are caused by the motion of the building. One way of testing the reliability of our measurement consists of comparing the measurements at two different locations of the building but at the same height. The displacements time series are different at the two locations; but, their frequencies of vibration should be the same if they are caused by the motion of the building.

For that purpose, we monitored the behavior of (23) located at the same height as (13), but not in the same column, as shown in Figure 17. Figure 21 represents the horizontal displacement of (13) and (23), and Figure 22 represents their vertical displacements. Figure 23 displays the normalized power spectral density of the horizontal displacements of (13) and (23), which seems to be similar. Likewise, Figure 20 displays the normalized power spectral density of the vertical displacements of (13) and (23); in both cases the resonant frequencies are identical.

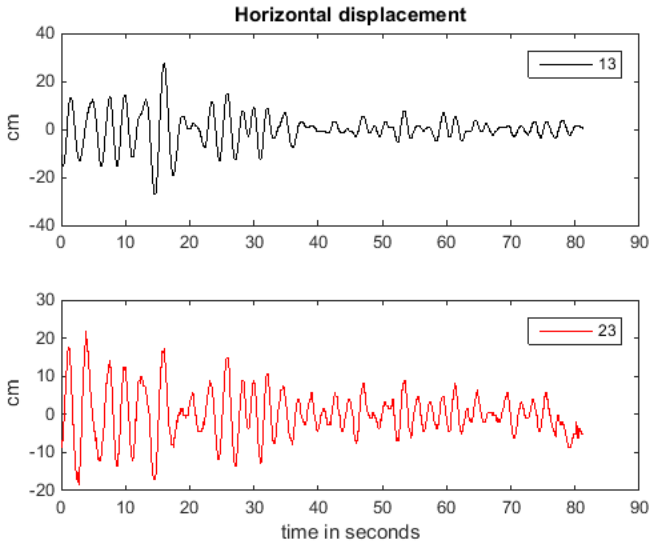


Figure 20 : Horizontal displacement of (13) and (23).

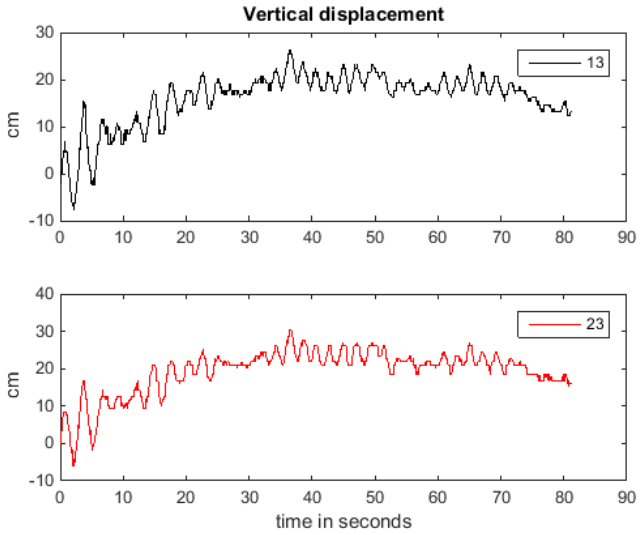


Figure 21: Vertical displacement of (13) and (23).

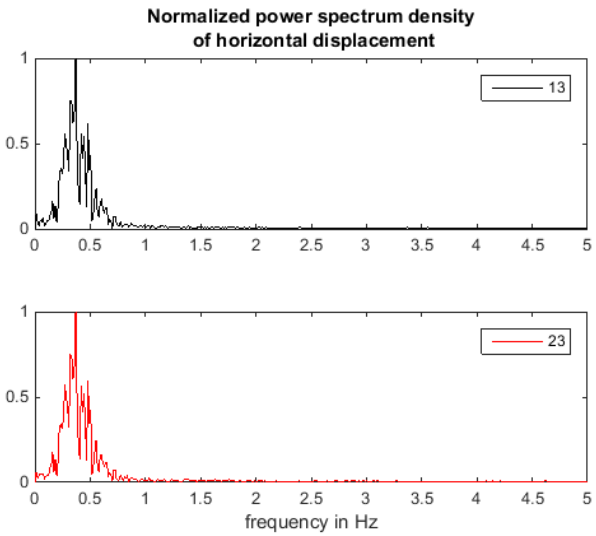


Figure 22: Normalized power spectrum density of horizontal displacements of (13) and (23).

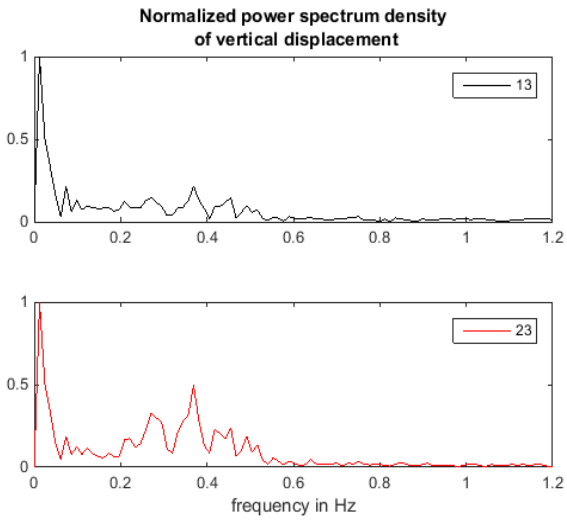


Figure 23: Normalized power spectrum density of vertical displacements of (13) and (23).

5.3 London Millennium Bridge

We intend to display the potential of our technique in measuring the physical quantities required for a thorough structural analysis of bridges. Nevertheless, a bridge structural analysis per say is not the prime objective of this section. According to the American Road and Transportation Builders Association (ARTBA), nearly 10% of the 609,539 bridges in the US are considered structurally deficient and need repairs. In this section, we look at the Millennium Bridge opened on 10 June 2000 as London's first new Thames crossing in more than 100 years. Tens of thousands of people had been crossing the bridge on its opening day. Suddenly, the bridge developed a very worrying and obvious wobble, in spite that its structure was designed to take the weight (Millennium Bridge). Since that event, the bridge has been nicknamed the "Wobbly Bridge".

Figure 24: View of the Millennium London Bridge along its suspension cables, (Wibbly Wobbly Millennium Bridge in London).

We downloaded from YouTube a film displaying the bridge just after its opening to the public, (Wibbly Wobbly Millennium Bridge in London). The video has a resolution of 2 megapixels and a frame rate of 29.97 images per second.

Figure 25 represents the vertical and horizontal displacements of one of the bridge's suspension cables, which are 12 cm in diameter according to (Millennium Bridge); Figure 26 represents the normalized power spectral densities of the same displacements.

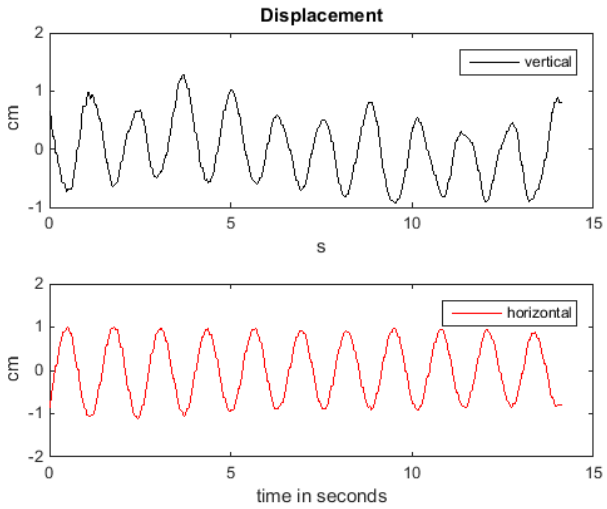


Figure 25: Vertical and horizontal displacements of one of the bridge's suspension cables.

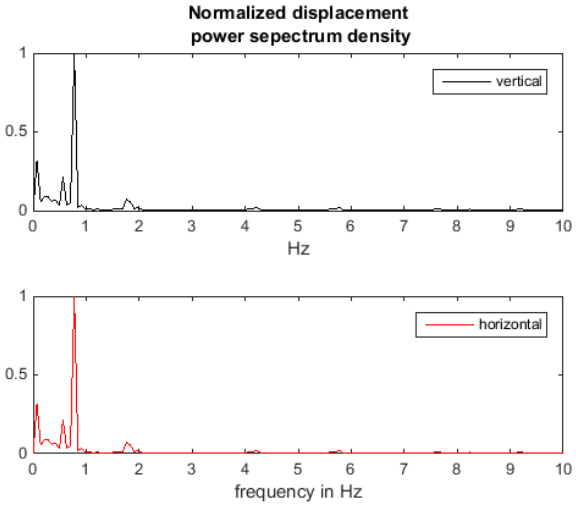


Figure 26: Normalized power spectrum densities of vertical and horizontal displacements.

5.4 Boeing 777

Analogous to the preceding study, we apply our technique to the introductory study of airplanes. We downloaded a video pertaining to a commercial airplane's wing from YouTube with a resolution of 2 megapixels and a frame rate of 29.97 images per second, (Boeing 777 Wing Flex Inflight Turbulence).

The description claims that the video pertains to a Boeing 777 of United Airlines flying over the coast of Japan at an altitude of 32,000 feet. Boeing 777 is the world's largest twin-engine jet airliners, and its turbofan engines have the largest diameter of any aircraft. Figure 27 is a snapshot of the video that we analyzed.

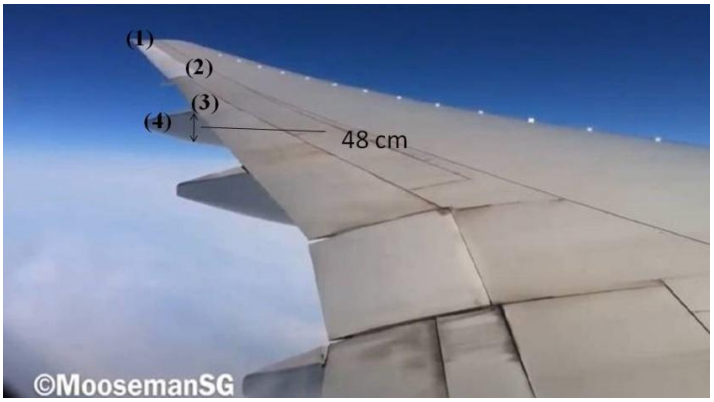


Figure 27: View of the airplane. In black are the regions of interest. From Figure 28 , we deduce the height of the canoe-shaped aircraft-fairing

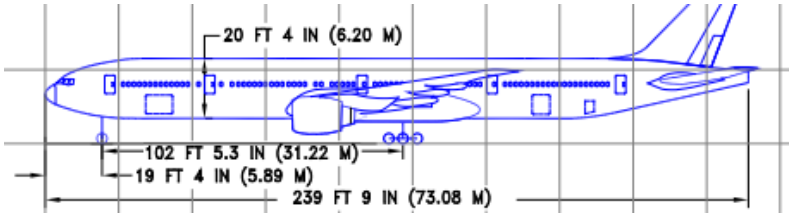


Figure 28: Accurate reproduction of the airplane, (Boeing Commercial Airplanes, 2015). Since the fuselage height is known, we deduce the height of the side of the canoe-shaped aircraft-fairing visible in the video footage.

Because of the field of view, the aircraft engine, which is situated below the wing, cannot be seen in the video.

We were not able to find a stationary object in the field of view that we could use as a reference to test the motion of the camera, consequently we cannot know whether the camera that took the movie was stationary relative to the fuselage of the airplane. In order to measure the motion of the wing, we selected several regions that are specified in Figure 27. However, we have to take into account that the image scale is not a constant; for example, it is different from at the end of the wing to what it is in the front view of the image. We mainly have to know the scales along

the direction perpendicular to the plane, since we are interested in measuring the out-of-plane motions of the targets. In addition, the variations in the scale of the image seem to be anisotropic. Indeed, a pixel in the direction of the plane of the wing seem to represent a different physical length than along the direction perpendicular to the plane of the wing

Though we measured the displacements of the four targets shown in Figure 27 , we can estimate their meanings in term of unit of length in the metric system for targets (3) and (4) only; indeed, we observe that targets (3) and (4) are in the vicinity of a canoe-shaped aircraft-fairing, whose function is to cover for gap between parts of the aircraft to reduce drag. We infer the vertical scale at targets (3) and (4) by estimating the vertical length of the structure. We estimate the height of the visible side of the aircraft-fairing as displayed in the video by using the schematic shown in Figure 28, which we believe being an accurate reproduction of the airplane. According to the constructor specifications, (Boeing Commercial Airplanes, 2015) , the fuselage is 6.2 m in height, which leads us to estimate the representation scale. Assuming that the vertical scale is approximately the same at any point in the schematic image, we guess the dimension of the aircraft-fairing to be roughly 0.48 m tall. We calculate the out-of-plane motions of (3) and (4) using the scale, which we estimated, that has been used to display the aircraft-fairing, Figure 29.

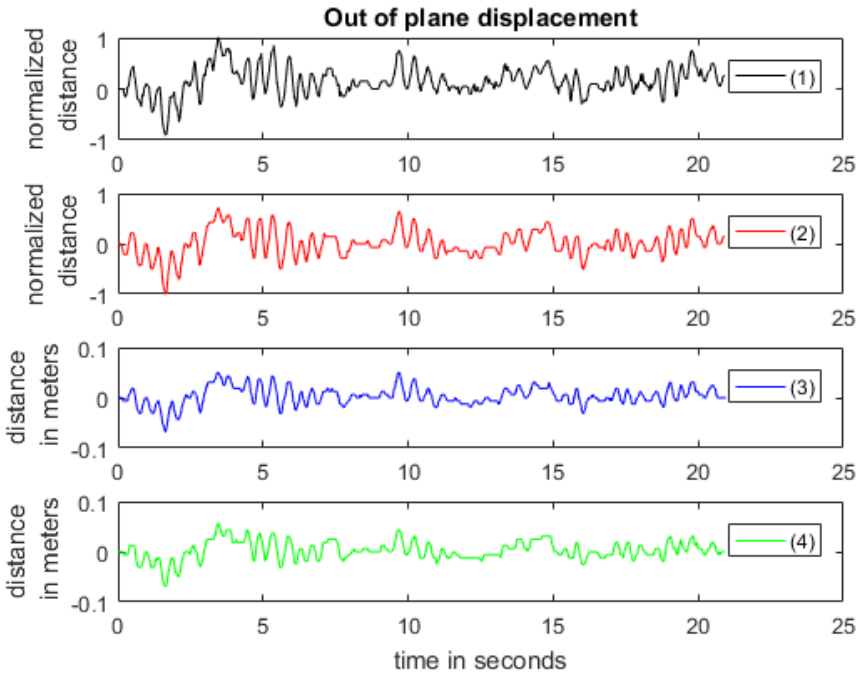


Figure 29: Out-of-plane displacement of each target. The displacements of (1) and (2) are normalized.

During the testing, we observed that the motion was mostly out of the plane of the wing at the tip-point of the wing. But, as the target that we measured the motion was getting closer to the cabin, the out-of-plane motion was attenuating, whereas the in-plane motion was increasing.

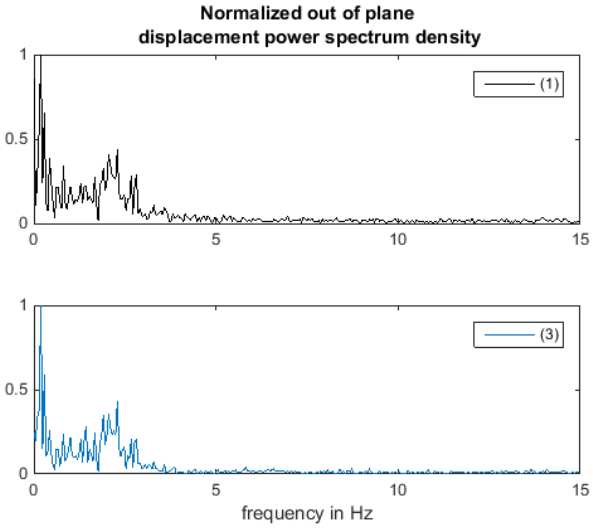


Figure 30: Normalized out-of-plane displacement power spectrum density of (1) and (3).

Moreover, it seems that the turbulence causes a quasi-periodic behavior of the wing, which induces pseudo-periodic movement characterized by the main peak in the power spectral densities, which is different from any of the structure resonances happening at higher frequencies, Figure 30.

5.5 NASA space shuttle main engines

In this section, we are interested in looking at the critical moment preceding the liftoff of NASA space shuttle using a video from YouTube, (Part 1: Space shuttle launches high-speed video camera slow motion views). It is claimed that the uploaded video was recorded using a high-speed camera, which typically has a recording rate of several hundred frames per second. Such high-frequency recording data is suitable for spotting high-speed events.

To find out the video's displaying speed, we used the LED display located on the right side of the screen, which uses a Universal Time Coordinated (UTC) time code format, to determine that 1 second in the video corresponds to 59 milliseconds in reality. However, the YouTube video exhibits a rate of 29.97 images per second, which means that one frame should correspond to 2 milliseconds in reality; that is, a frame rate of 500 Hz. That is different than the camera's speed of 400 Hz as stated by the video voiceband. In addition, it is not granted that the video we are working with is not missing frames from its original version even though the video is displayed in slow motion.

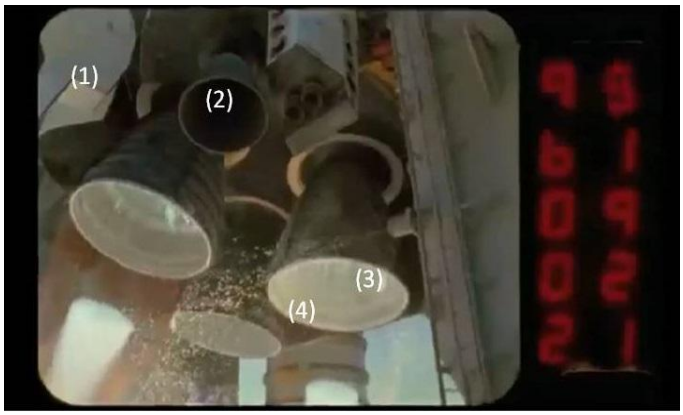


Figure 31: View of the rockets exhausts. In white are the objects we monitored.

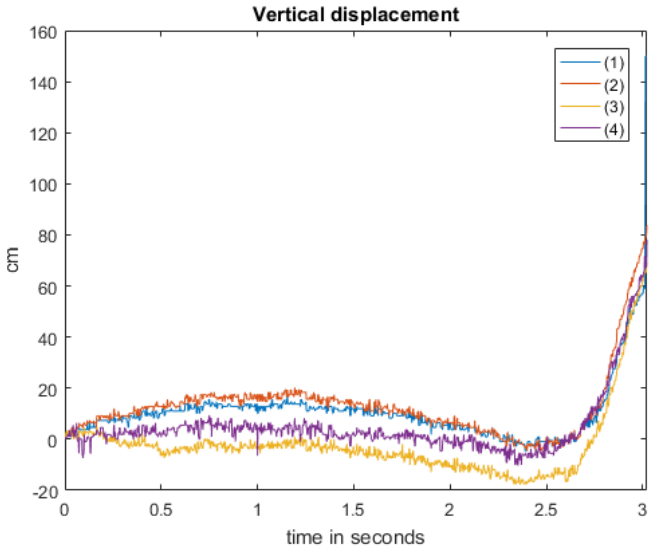


Figure 32: Vertical displacement of each target.

To approximately estimate the order of magnitude of the scaling in the image, we alluded to the dimension of the solid rocket nozzle extension, which we proximate similar to the 3.71 m long diameter solid rocket booster, (NSTS Shuttle Reference Manual (1988), 1988); but deprived of further information pertaining to the orientation and position of the camera relative to the scenery, we assume them constants during the course of the recording. Neverthless, this is a coarse approximation since the motions, which are visible to a human eye , happen in a field of view that has perspective distortion. Indeed, the voiceband of the video states that a lens with 10 mm focal length is used; that is, it is a wide-angle lens. A result of using a wide-angle lens, as we saw earlier, is a great perspective distortion especially when the camera is not aligned perpendicularly to the subject, as in this video.

So, as a target on the rocket moves in the visible scenery, the ratio of the length of the displacement as perceived in the image to the length of the real displacement changes; but deprived of further information, we neglect variations in scale.

Figure 32 displays the vertical displacements of the targets. We observe from the moment, which seems corresponding to the release of the rockets, a decrease in intensity of the vibrations.

Figure 33 displays the normalized vertical displacement power spectral density of the targets. With the naked eye, it seems that the normalized power spectrum density of (2), (3), and (4) are very similar, while the one of (1) exhibits slight slower decay as frequency increases.

Another interesting observation pertains to what might be caused by the modal vibrations of the rocket nozzle extension. Indeed, we observe a time varying displacement between two different regions of the same rocket nozzle extension, as if it experiences structural modal vibrations. Figure 34 displays the vertical displacement of (3) relative to (4), whose power spectral density exhibits a peak at 0.6 ± 0.33 Hz, as show in Figure 35. However, we need a longer time series data (at least 10 times longer) to resolve the uncertainty in the frequency and to have more confidence that the peak we observe is the result of a real physical phenomenon and not some other artifacts.

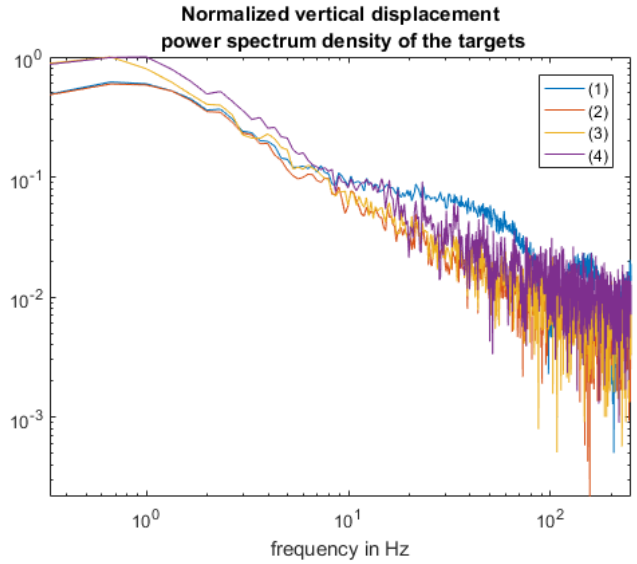


Figure 33: Normalized vertical displacement power spectrum density of the targets; the time series data is shown in Figure 32.

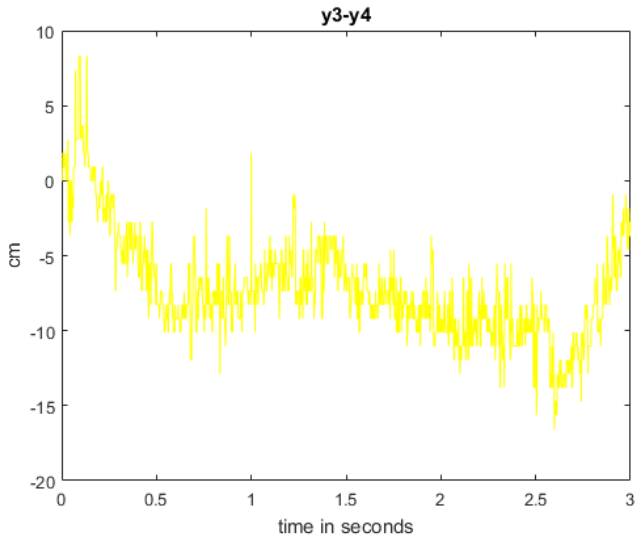


Figure 34 Vertical displacement of (3) relative to (4), y3-y4. It exhibits a resonance frequency at 0.66 Hz, as shown in Figure 35.

Without more technical information about the characteristic and the environment of the experiment, we cannot state that our observations have not been caused or at the least not been affected by other factors: heat (shimmering), a flow of particles, and etc., coming out of the

engine and causing light scattering. Nevertheless; it has been reported in the past that the space shuttle booster rocket had experienced structural modal vibrations, (Mason, Morstadtt, Cannon, Gro, & Nielson, 2004) ; that is, a pattern of motion in which all parts of the system move sinusoidally with the same frequency and with a fixed phase relation. The most general motion of such a system is a superposition of its normal modes; the modes are normal in the sense that they can move independently, that is to say, that an excitation of one mode will not excite a different modeb . So, the observed frequency is perhaps the frequency of a normal mode of resonance.

Therefore, we convoluted the measurement data with $\frac{1}{2\pi}\, e^{-0.04t} c\,\mathrm{os}(2\pi 0.66t)\ (t > 0)$, which corresponds to apply a temporal filter at a frequency of 0.66 Hz with a decay rate of 4% to the data; as it is displayed in Figure 36. This operation equals to multiply the Fourier transform of the displacement with a Cauchy-Lorentz distribution centered at 0.66 Hz, and then, take the inverse Fourier transform of the result.

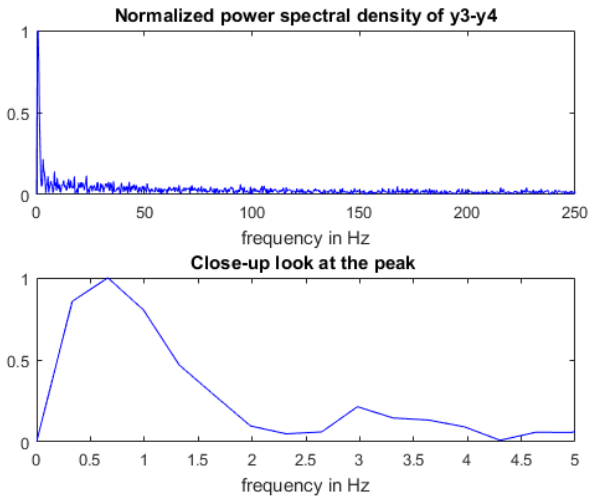


Figure 35: Normalized power spectral density of y3-y4; the time series data is shown in Figure 34 . It exhibits a peak at 0.66 ± 0.33 Hz.

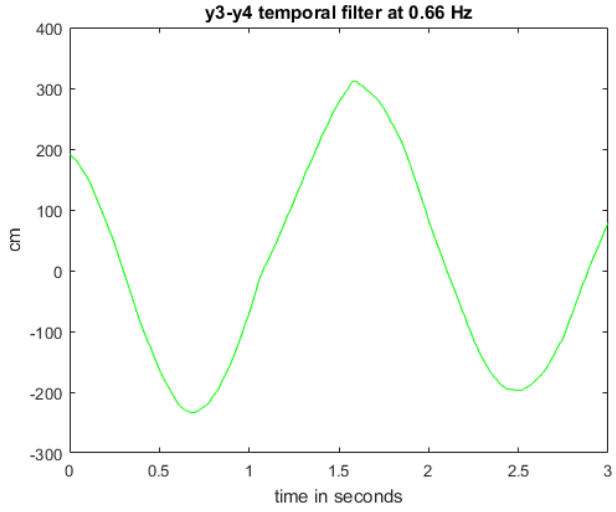


Figure 36: Vertical displacement of (3) relative to (4) when a temporal filter at a frequency of 0.66 Hz with a decay rate of 4% is applied.

Physically speaking, let's characterize the system as a linear time-invariant (LTI) mechanical oscillator. Let's assume that the measured relative displacement time series is the impulse response of the system. The convolution that we performed is similar than exciting the LTI

system at the input with a signal of a frequency of 0.66 Hz. If that corresponds to a resonant frequency of the system, then the output, which is the result of the convolution, will be large. That is why the result has such large amplitudes, i.e. one order of magnitude larger than the amplitudes of the original signal; the size of the output signal depends on the Q of the system at that frequency.

6 Conclusion

We looked at the feasibility of measuring structural deformations using optical videos recorded by cameras located away from the targets, (Taghavi Larigani & Heaton).

We displayed the results of measurements that we performed on 1) JPL/NASA principal building, 3) the Millennium London Bridge nicknamed the "Wobbly Bridge", 4) a United Boeing 777, and 5) NASA space shuttle rockets during launch. So far, our tests are encouraging.

Our technique consists of recording a video of unapproachable or inaccessible edifices and structures, which otherwise would have been instrumented with a series of accelerometers and seismometers, by selecting regions of interest (targets) to measure their displacements.

Since image scaling and perspective distortions remain constant in our application, we developed a straightforward algorithm relying on a normalized two-dimensional cross-correlation. In addition to being simple, this technique has the advantage of being invariant to global exposure and luminosity variation taking place during the course of the film. Hence, it is relatively robust and easy to implement. We were able to reduce the implications of the algorithm caused by its rotational-depending characteristic to an inconsequential importance, nevertheless using that same sensitivity relative to rotations, we are in the process of developing a tool to measure the spatial rotation of a target from two-dimensional video images of it. Additionally, we intend to

complement our Lagrangian approach to the problem with a Eulerian approach, where we will analyze the temporal inherent evolution of each target as well.

High-quality commercial-grade imaging systems (lens apparatus) are suitable to perform measurement of sub-mm resolution provided that the imaging systems are diffracted-limited. In regards to commercial grade digital cameras, they provide measurement strain up to few parts per ten-thousand (i.e. a dynamic range of four orders of magnitude). We are currently investigating to see if with these numbers we can use commercial-grade hardware to monitor reliably civil structures deformations.

We learned that downloading videos from the World Wide Web have limitations in quality, data rate and the segment of the video that is useful for our analysis is ,many cases, short.

If our approach proves viable, it could be transformative for the field of earthquake engineering, and structural health monitoring. Hence, we are contemplating the possibility of utilizing our method for surveying buildings and other civil structures in high seismic risk urban agglomerations.

In parallel, the same technique could be applied for real-time structural health monitoring of

- civil structures,
- nuclear plants,
- oil and gas infrastructures,
- rail & road networks,
- aircraft,
- spacecraft,
- etc.,

by simply analyzing the structure-facing camera recorded data.

Beyond the technical appeal, its seemingly relative low-cost deployment and not requiring the targeted structure owner's approval could be of interest for potential users. Other fields of science and engineering could as well find interest in utilizing it.

7 Acknowledgments

The authors are thankful to Anthony Massari for providing ETABS generated data pertaining to JPL/NASA building 180 dynamics.

Shervin Taghavi Larigani is grateful to Mr Micheal David Donohoe (STL-Scientific), and thanks Dr Jakob Vanzyl (Caltech & NASA/JPL), Prof Michael Hoffmann (Caltech), Prof Mark Simons (Caltech), Prof Luis Rivera (Caltech & Institut de Physique du Globe de Strasbourg), Prof James Rice (Harvard University), Mr B. Afshar, Dr Renaud Goullioud (JPL/NASA), Mr A. Gerber (JPL/NASA), Colonel B. Cox (USAF ret & JPL/NASA), Prof Christopher Boxe (City University of New York), Prof Guruswami Ravichandran (Caltech), Dr Thomas Ader (Caltech), Mr J. Kelin, Dr Andrea Donnellan (JPL/NASA), Major F. Nassirkhani (IIAF pilot ret), Prof Andrew Ingersoll (Caltech), Prof Yuk Yung (Caltech), Dr Ramses Mourhatch (JPL/NASA), Prof John Baldeschwieler (Caltech emeritus), Dr Ken Pickar (Caltech), Mr R. Stuchter (Programmed Scientific Instruments) and last but not least Mr Sang Chung (JPL/NASA) for their fruitful conversations and help.

8 Bibliography

50 years of talking to space a Nasa social for the deep space network. (n.d.). (universetoday) Retrieved from http://www.universetoday.com/

Alvarez, L., Gomez, L., & Sendra, J. R. (n.d.). Algebraic Lens Distortion Model Estimation. *Image Processing On Line, 1*, 2010.

Bay, H., Tuytelaars, T., & Van Gool, L. (2006). SURF:Speeded Up Robust Features. *European Conference on Computer Vision.* Graz, Austria.

Beauchemin, S. S., & Barron, J. L. (1995). The computation of optical flow. *ACM Computing Surveys (CSUR)*, 433-466.

Betzig, E., & Trautman, J. (n.d.). Near-field optics:microscopy, spectroscopy, and surface modification beyond the diffraction limit. *Science*(257), 189-195.

Boeing 777 Wing Flex Inflight Turbulence. (n.d.). (YouTube) Retrieved from https://www.youtube.com/watch?v=GzvA9tHNTyg

Boeing Commercial Airplanes. (2015). *777-200LR/-300ER/-Freighter Airplane Characteristics for Airport Planning.* BOEING.

Born, M., & Wolf, E. (n.d.). *Principles of Optics.* Cambridge Univ.Press.

Browne, M. (n.d.). *Photography Tips Techniques and Tutorials*. Retrieved from YouTube: https://www.youtube.com/channel/UCs4S07E5NC0R7pVuzBJ0e8Q

Bucher, C. (n.d.). *Lighting Photo Workshop.* John Wiley and Sons.

California Institute of Technology. (n.d.). (Wikipedia) Retrieved from https://en.wikipedia.org/wiki/California_Institute_of_Technology

Cheng, M., & Heaton, T. (2015). Simulating building motions using ratios of the building's natural frequencies and Timoshenko Beam Model. *Earthquake Spectra*, 403-420.

Chong, A. K., Al-Baghdadi, J. A., & Alshadli, D. (2014). High Definition Video Cameras for Measuring Movement of Vibrating Bridge Structure. *International Conference on Vibration and Vibro-acoustics* . Harbin, China.

Clayton, R. W., Heaton, T., Chandy, M., Krause, A., Kohler, M., Bunn, J., . . . Aivazis, M. (2011). Community Seismic Network. *Annals of Geophysics*.

Cox, I., & Sheppard, C. (1986). Information capacity and resolution in an optical system. *Journal of the Optical Society of America A, 3*, 1152–1158.

Cozzolino, P. (n.d.). *Millennium Bridge, London*. (National Geographic) Retrieved from http://travel.nationalgeographic.com/travel/countries/your-england-photos#/millennium-bridge-london_33512_600x450.jpg

Dalal, N., & Triggs, B. (2005). Histograms of Oriented Gradients for Human Detection. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR).* San Diego.

Di Francia, T. (1955). Resolving power and information. *Journal of Optical Society of America, 45*, 497–501.

Donner, N. (2014, June 20). *Most Important Factors for Selecting a DSLR Memory Card*. Retrieved from http://www.craftsy.com/blog/2014/06/digital-photography-memory-cards/

Donohoe, M. D. (n.d.). Oral Communication.

E-defense Dynamic Test of Seismic Isolated Building by Sannomaru Eq. (n.d.). (YouTube) Retrieved from https://www.youtube.com/watch?v=qIwFz6cxcOE

Farsjiu, S., Robinson, D., Elad, M., & Milanfar, P. (2004). Fast and Robust Multi-frame Super-resolution. *IEEE Transactions on Image Processing, 13*(10), 1327–1344.

Fleet, D. J., & Weiss, Y. (n.d.). Optical flow estimation. In *Handbook of Mathematical Models in Computer Vision* (pp. 237-257). Springer US.

Fox, M. (2007). *Quantum Optics Oxford.*

Gaskill, J. D. (1978). *Linear Systems, Fourier Transforms, and Optics.* Wiley-interscience.

Goodman, J. W. (2004). *Introduction to Fourier Optics (Third Edition).* Roberts & Company Publishers.

Gustaffsson, M. (2000). Surpassing the lateral resolution limit by a factor of two using structured illumination microscopy. *Journal of Microscopy, 198*, 82-87.

Harris, J. (1964). Resolving power and decision making. *Journal of Optical Society of America*(54), 606–611.

Histogram of oriented gradients. (n.d.). (Wikipedia) Retrieved from https://en.wikipedia.org/

Image Correction for Lens Distortion. (n.d.). (MathWorks) Retrieved from http://www.mathworks.com/help/vision/ref/undistortimage.html

Lowe, D. G. (1999). Object Recognition from Local Scale-Invariant Features. *Proceeding of the International Conference on Computer Vision.* Corfu.

Lui, C., Torralba, A., Freeman, W. T., Durand, F., & Adelson, E. H. (2005). Motion magnification. *ACM Trans. Graph.*(24), 519-526.

Lukosz, W. (1966). Optical systems with resolving power exceeding the classical limit. *The Journal of the Optical Society of America, 56*, 1463–1472.

Martin, J. (2004). *Digital Photography Outdoors: A Field Guide for Travel and Adventure Photographers The.* The Mountaineers Books.

Mason, D. R., Morstadtt, R. A., Cannon, S. M., Gro, E. G., & Nielson, D. B. (2004). Pressure Oscillations and Structural Vibrations in Space Shuttle RSRM and ETM-3 Motors. *40th AIAA/ASME/SAE/ASEE Joint Propulsion Conference and Exhibit.* Fort Lauderdale, Florida.

Massari, A. (n.d.). . Retrieved from http://www.massari.caltech.edu/

Millennium Bridge. (n.d.). (BBC) Retrieved from http://news.bbc.co.uk/hi/english/static/in_depth/uk/2000/millennium_bridge/default.stm.

(1988). *NSTS Shuttle Reference Manual (1988).* NASA.

Panasonic. (n.d.). *About SDXC*. Retrieved from Panasonic: http://panasonic.net/avc/sdcard/information/sdxc.html

Part 1: Space shuttle launches high-speed video camera slow motion views. (n.d.). (YouTube) Retrieved from https://www.youtube.com/watch?v=wlz5u1OBe_c

Poot, D., Jeurissen, B., Bastiaensen, Y., Veraart, J., Van Hecke, W., Parizel, P. M., & Sijbers, J. (2012). Super-Resolution for Multislice Diffusion Tensor Imaging. *Magnetic Resonance in Medicine*.

Schuon, S., Theobalt, C., Davis, J., & Thrun, S. (2009). LidarBoost: Depth Superresolution for ToF 3D Shape Scanning. *IEEE Proceedings of IEEE Computer Vision and Pattern Recognition.*

Shaw, P. J., & Rawlins, D. J. (1991). The point-spread function of a confocal microscope: its measurement and use in deconvolution of 3-D data. *Journal of Microscopy*, *163*(2), 151–165.

Simpkins, J., & Stevenson, R. (2012). An Introduction to Super-Resolution Imaging." Mathematical Optics: Classical, Quantum, and Computational Methods, Ed. V. Lakshminarayanan, M. Calvo, and T. Alieva. CRC Press, 2012. *Mathematical Optics: Classical, Quantum, and Computational Methods CRC Press*, 539-564.

Smith, & Warren. (2005). *Modern Lens Design.* McGraw-Hill.

Sony. (n.d.). *HX300 Camera with 50x Optical Zoom*. Retrieved from Sony-Electronics: http://www.sony.com/electronics/cyber-shot-compact-cameras/dsc-hx300

Starr, C., Evers, C., & Starr, L. (2005). *Biology: Concepts and Applications* . Thomson Brooks/Cole.

(November 2006). *Statistical Loads Data for the Boeing 777-200ER Aircraft in Commercial Operations* . Washington, DC 20591: U.S. Department of Transportation, Federal Aviation Administration, Office of Aviation Research and Development .

Taghavi Larigani, S., & Heaton, T. H. (n.d.). *Patent pending, STL-Scientific & Caltech.*

Westheimer, G. (2012). Optical superresolution and visual hyperacuity. Prog Retin Eye Res. *Progress in Retinal and Eye Research, 5*(31), 467-80.

Wibbly Wobbly Millennium Bridge in London. (n.d.). (YouTube) Retrieved from https://www.youtube.com/watch?v=7GuRPlWnrto

Wu, H.-Y., Rubinstein, M., Shih, E., Guttag, J. V., Durand, F., & Freeman, W. T. (2012). Eulerian video magnification for revealing subtle changes in the world . *Association for Computing Machinery*.

Zalevsky, Z., & Mendlovic, D. (n.d.). *Optical Superresolution.* Springer.

Zhou, X., Xiab, Y., Weib, Z., & Wu, Q. (2012). A videogrammetric technique for measuring the vibration displacement of stay cables. *Geo-spatial Information Science, 2*(15), 135-141.

Introduction

- **Optical flow**

  The optical flow methods try to calculate the relative motion between two consecutive image frames by supposing that the motion is a continuous and somehow a steady stream, hence the name of optical flow (Beauchemin & Barron, 1995) . They consist of pairing each pixel coordinates in the first image to its coordinate position in the second image; that is, they use local Taylor series of the image signal. In other words, they use partial derivatives with respect to the spatial and temporal coordinates; that is why they are also referred as differential methods. They consider that a pixel, at position (x,y), in time t, and with the intensity $I(x, y, t)$, moves by Δx and Δy at the next iteration $t + \Delta t$, such

$$I(x, y, t) = I(x + \Delta x, y + \Delta y, t + \Delta t).$$

That is called the image brightness constancy equation, and assumes the scene does not change intensity. Its Taylor series linear approximations, which can only be used if the motion is small (corresponding to small-perturbation approximations), leads to

$$I(x + \Delta x, y + \Delta y, t + \Delta t) \approx I(x, y, t) + \frac{\partial I}{\partial x}\Delta x + \frac{\partial I}{\partial y}\Delta y + \frac{\partial I}{\partial t}\Delta t.$$

The combination of the two previous equations leads to,

$$\frac{\partial I}{\partial x}\Delta x + \frac{\partial I}{\partial y}\Delta y + \frac{\partial I}{\partial t}\Delta t \approx 0.$$

Dividing the previous equation with Δt results in,

$$\frac{\partial I}{\partial x}V_x + \frac{\partial I}{\partial y}V_y + \frac{\partial I}{\partial t} \approx 0,$$

where $V_x = \frac{\Delta x}{\Delta t}$, $V_y = \frac{\Delta y}{\Delta t}$ are the velocity (or optical flow) of $I(x, y, t)$ along the x and y coordinates, and $\frac{\partial I}{\partial x}$, $\frac{\partial I}{\partial y}$, $\frac{\partial I}{\partial t}$ are the derivatives of the image at (x,y) and time t. This equation only considers a single point as it changes with time. We can rewrite the previous equation in its matrix form,

$$(\nabla \boldsymbol{I}^{\boldsymbol{T}}).\boldsymbol{V} \approx -\frac{\partial I}{\partial t} ,$$

where $\boldsymbol{I} = \begin{pmatrix} I_x \\ I_Y \end{pmatrix}$, and $\boldsymbol{V} = \begin{pmatrix} V_x \\ V_Y \end{pmatrix}$ are two-dimensional column vectors, and ∇ is the nabla operator. Since this equation is for a single point, we have an equation with two unknowns. So, additional information is required to recover the motion. Otherwise, an aliasing problem takes place where a solution satisfying the equation does not necessarily correspond to the true motion underlined in the image frames. That is known as the aperture problem of optical flow (Fleet & Weiss).

Let's suppose that we are interested in detecting the motion of an object that is constituted of parts that have each orientation and motion distinct than another one part. Let's also suppose that the object occupies a large fraction of the entire visual field and we only have a single-cell motion sensor to detect the object's motion.

As the object moves relative to the detector, the intensity of the light captured by the detector changes. Nevertheless, that information is insufficient to infer the direction of the motion; hence, the motion-sensor would fail to recognize the true motion of the object. The scenario that we just described illustrates the aperture problem of optical flow. Hence, all optical flow methods introduce additional constraints for estimating the true flow. That is why it is assumed, when feasible, that the motion remains constant over a window as an example.

In our approach, we don't rely on brightness constancy. In addition, the displacements we measure are not necessarily homogenous over the volume of the structure that we test, nor are they necessarily linear.

As to the causes of the motion we measure, they are often high-energy phenomena, which induce perturbations that cannot be approximated as small. In the example of a building, deformations happen because of irreversible changes of shape.

- **Autonomous station**

One could imagine a full autonomous station-unit consisting of one or several cameras connected to a highly stable physical mounting. The physical apparatus could include a carrier-phase enhancement GPS receiver unit (CPGPS) that would calculate the absolute location of the station, and thus would measure the station motions, which would be subtracted from the motions of the targets measured by the cameras to provide the absolute motions of the targets. Current CPGPS measurements have resolutions of few cm and better.

Accelerometers are suited to measure motions of frequencies higher than 1 Hz. On the other hand, GPS measurement noise converges to the Ultra Stable Oscillators of the GPS constellation satellites over time. That signifies that the GPS is the best fit to pick up very low-frequency motion, thus complementing an accelerometer.

All the electronics in the unit would be interfaced to a microcontroller/computer that would act as the system instrumentation platform, and would process and send the data wirelessly or via the internet to the users.

In addition, the unit could be powered by a solar panel during the day, which would charge a battery that could be used to power the system during the night

Hardware and Measurement Characteristics

- Since our targets are relatively distant from the camera, a tiny change in the distance separating the camera's frame from the focal plane of the lens apparatus can bring the image into focus, thus capturing a sharp video of the target should not be an obstacle as long the image of the target on the camera's frame is larger than the size of the circle of confusion, which is an optical spot caused by a cone of light rays from a lens coming to focus when imaging a point source (even at best focus, a point is imaged as a spot rather than a point). Indeed, the thin- lens approximation leads to:

$$\frac{1}{L_z}+\frac{1}{L_i}=\frac{1}{f}, \qquad 3$$

where L_z is the range distance between the center of the lens and the target, f and L_i are in the order of magnitude of millimeters while L_z is in the order of magnitude of tens of meters, thus 1/f and $1/L_i$ are four order of magnitudes larger than $1/L_z$. Accordingly, the required adjustment of L_i (the camera's frame) to get a sharp image is in the order of magnitude of millimeters at most.

- **Imaging system diffraction limit**

  According to the Huygens–Fresnel principle, diffraction is described as the interference of waves (light in our case). As mentioned in the introduction, the response of an imaging system to a point source due to diffraction is not a point source but a pattern, whose central figure is the result of the spreading of a point in space, hence the name of Point Spread Function (PSF). A more general term for the PSF is a system's impulse response, the PSF being the impulse response of a focused optical system, (Shaw & Rawlins, 1991)(Goodman, 2004).The diffraction pattern resulting from a uniformly-illuminated perfect lens with a circular aperture has a bright region in the center, known as the Airy-disk, which together with the series of concentric bright rings around is called the Airy-pattern. The theoretical estimation of the diffraction induced resolution of the measurement consists of determining the characteristics of the Airy-disk; thus, discerning consecutive light-intensity maximum and minimums on the Airy-pattern.

  One of the standards of detection, namely the Rayleigh criterion, specifies that on the line between the center of one point and the next, the contrast between the maximum and minimum intensity is at least 26% lower than the maximum, (Gaskill, 1978). This corresponds to the overlap of one airy disk on the first dark ring and the other, where the minimum distance between resolvable points is defined as:

$$r_{min} = \frac{0.61\ \lambda}{N_A}, \qquad 4$$

  where N_A is the numerical aperture, which is defined as

$$N_A = 2\, n \sin(\theta), \qquad 5$$

where λ is the wavelength of light, n is the index of refraction of the media surrounding the radiating points, θ is the half angle of the maximum cone of light that can enter or exit the lens as illustrated in Figure 6.

In photography the concept of f-number (denoted here as N_f) embodies similar idea as the numerical aperture; it accurately describes the light-gathering ability of a lens only for

objects of an infinite distance away (Martin, 2004)(Bucher). This constraint is typically suitable for photography, where objects are usually not extremely close to the camera, relative to the distance between the lens and the film.
The f-number (sometimes called focal ratio, f-ratio, f-stop, or relative aperture) of an optical system is the ratio of the lens's focal length to the diameter of the entrance pupil, and is given by:

$$N_f = \frac{f}{A_{eff}} \qquad 6$$

where f is the focal length, and A_{eff} is the diameter of the entrance pupil (effective aperture). θ is a function of f, A_{eff} , and is expressed as,

$$\theta = \arctan\left(\frac{A_{eff}}{2f}\right). \qquad 7$$

Equations 7 and 6 lead to express θ in function of N_f ,

$$\theta = \arctan\left(\frac{1}{2N_f}\right), \qquad 8$$

Since

$$\sin(\arctan(x)) = \frac{x}{\sqrt{1+x^2}}, \qquad 9$$

r_{min} could be defined in function of N_f using equations (4-9) as,

$$r_{\min} = 1.22\lambda \frac{\sqrt{1+\left(\frac{1}{2N_f}\right)^2}}{n\left(\frac{1}{2N_f}\right)},$$

where λ is the average visible light wavelength that a typical human eye responds to, i.e. from about 390 to 700 nm, it is thus estimated to 550 nm, (Starr, Evers, & Starr, 2005); and the index of refraction of air being approximated to 1.

Most modern lenses use a standard f-stop scale, which is an approximate geometric sequence of numbers that corresponds to the sequence of the powers of the square root of 2, which signifies that the f-number being equal to 1, 1.4, 2, 2.8, 4, 5.6, 8, 11, 16, 22, 32, 45, 64, 90, 128, etc. However, most camera lenses have f-number varying from 4 to 32. Shorter f-numbers are typical of telescopes. Assuming a pixel size of 6.5x6.5 um^2 we deduce the plot in Figure 37, which represents the diffraction induced response of imaging systems (defined by their f-numbers) to a point source under the plane-wave approximation.

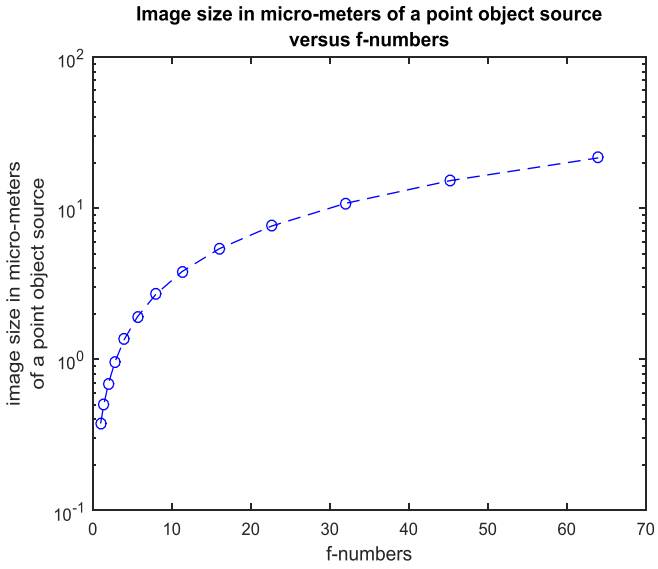


Figure 37: Image size in micro-meters of a point object source versus f-numbers.

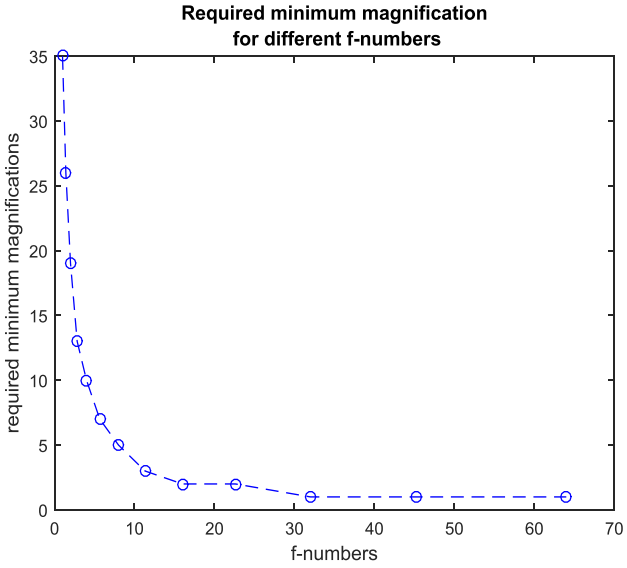


Figure 38: Required minimum magnification for different f-numbers.

In order to optimize the camera so that it reaches the resolution limit dictated by diffraction, the Airy-disk, representing the system's impulse response to a point source should be magnified such that its enlarged size becomes at least larger than two pixels length. Figure 38 describes the minimum magnification needed for each f-number assuming a pixel size of 6.5x6.5 um

Two adjacent points in the object give rise to two diffraction patterns. If the angular separation of the two points is significantly less than the Airy-disk angular radius, then the two points cannot be resolved in the image, but if their angular separation is much greater than this, distinct images of the two points are formed and they can, therefore, be resolved. Similarly, there exist arbitrary criterions such as "Rayleigh angular separation

criterion" that two points whose angular separation is equal to the Airy disk radius to first null can be considered to be resolved.

$$\theta_{\min} = 1.22 \frac{\lambda}{A_{eff}}.$$

It can be seen that the greater the diameter of the lens or its aperture, the greater the resolution. Astronomical telescopes have increasingly large lenses so they can 'see' ever finer detail in the stars.
In order to transliterate the angular resolution criterion in an expression more relevant to digital cameras, we express Aeff as a function of the f-number, which leads to an expression of the angular resolution of

$$\theta_{\min} = 1.22 \lambda \frac{N_{eff}}{f},$$

Figure 7 represents angular resolutions in radians versus f-numbers for the three types of lenses we have considered in this manuscript. Thereby, by multiplying the previous expression with the distance separating the objects from the camera we calculate the resolution of measuring displacement at the target.

- **Numerical application in the case of an entry-level camera, the Sony DSC-HX300**
  The Sony DSC-HX300 (Sony) entry-level digital camera, which only costs a few hundreds of US, has a high-quality imaging lens comprised of a varying effective focusing length that can be as large as 1200 mm, which corresponds to an angle of view that could be as small as of 20 mrad along the vertical direction and of 30 mrad along the horizontal direction. As stated by the manufacturer, the length of the diagonal of the sensor is 7.82 mm; that is 5.5 times smaller than the diagonal of a full frame size camera sensor. Therefore, we conclude that the sensor has a crop factor of 5.5 where the crop factor is defined as the ratio of the dimensions of the sensor to the dimensions of a full frame film format. We estimate the sensor being 6.54 mm large and 4.36 mm tall. The extent of the observable scenery that is distant of Lz along the direction of the optical axis of the camera as it is seen from the camera equals Lz multiplied by the camera's angle of view. We seek to know how far from the camera the target could be when one pixel in the camera's sensor represents a length of one cm at the scenery. Given that the camera has 20.4 megapixels, we estimate that a unit pixel represents a length of one cm at the field of view when the targets are at a distance of 1.84 km from the camera.

- **Camera stabilization**

The camera stabilization musst not induce error in the measurement that is equal or bigger than the resolution of the measurement. A wide range of professional photography, video and telescope mountings exists on the market that has demonstrated to be well-secured to the ground or a wall.

Image stabilization hardware corrects the errors that are induced by the lens yaw and pitch rotations by using a lens that floats in the plane perpendicular to the optical axis of the camera. A motion of the camera gear induces a relative displacement with the floating lens, which motion is detected by miniature piezoelectric sensors, whose output signals are treated by a high-speed computer, all built into the lens. Similarly, advanced camera sensors deploy a floating frame, whose motion relative to the camera gear is monitored.

If one is interested in measuring the absolute motion of the target (as opposed to its motion relative to the camera), one could deploy a series of accelerometers and gyroscopes that would measure the displacement and rotation of the camera relative to an inertial frame. By subtracting them from the motion of the target relative to the camera, one calculates the motion of the target relative to an inertial frame.

In addition to the mechanical calibration, we might be able to use to our benefit software-based image stabilization techniques. Most of these algorithms rely on canceling image-blurriness, which is induced by the image getting out of focus. The target moves either in the plane perpendicular to the camera's optical axis, which does not affect the sharpness of the captured image, or along the camera's optical axis. By examining equation (3), and referring to our earlier discussion, Lz is both four orders of magnitude larger than Li and the focal length of the lens; where Li and f are in the same order of magnitude. A hypothetical extreme motion of the monitored structure is at most in the orders of magnitude of meters (and in the worst-case) along the camera's optical-axis would still be within the depth of field, consequently, the image of the targets would still remain sharp. In contrast, yaw and pitch rotations of the camera blur the image (out of focus). Therefore, unblurring the image might cancel the effects of the camera's yaw and pitch rotations.

- **Change in light's intensity**

  As long as the target is visible, our algorithm is robust enough not to be affected by changes in light-intensity. However, if one is interested in sub-pixel measurement resolution then light-background-intensity changes would matter.

- **In-homogeneity in pixel dimensions**

  To evaluate the error that is induced in the measurement by the inhomogeneity in the size of the pixels, we look at the fabrication resolution of the photo-sensors, which is determined by the lithography technique used. Current dominant lithography technology uses deep ultraviolet (or DUV) light from excimer lasers with wavelengths of 248 and 193 nm, which allow minimum feature sizes down to 50 nm. That is less than 1% of the size of the pixel size we are considering in this analysis.

That is, inequality in the size of the pixels induces error that is an insignificant percent of the measurement resolution, thereby we neglect it in our analysis.

- **Lens distortion**
  Radial distortion is the most commonly encountered category of lens distortions. The software Matlab (developed by MathWorks) provides camera calibration tools, which once the camera has been calibrated, corrects an image for lens distortion, (Image Correction for Lens Distortion). Many similar algorithms are available online, (Alvarez, Gomez, & Sendra).

- **Change in air-index of refraction**
  A procedure of perceiving the ramification of such phenomena on our measurement consists of taking color pictures. A color picture is constructed by adding the outputs of red, green, and blue detectors in one image. Instead, we could analyze separately the output images. We would select a specific feature on the capture scenery and look at its locations on each output of the detectors. Since the index of refraction is a function of wavelength (color), the same pattern might have a slightly different location depending on the wavelength (color) of the detector. If their relative distances remain constant over the course of the experiment, it means the change in the index of refraction of air has not induced error in the measurement. In contrary, if their relative locations have changed over the period of the experiment, we could estimate the contribution of the air index of refraction in our error budget by measuring the changes over time.

  To understand the ramification of changes in the air index of refraction on our measurement, we could place a series of heaters at different locations of a structure and change in-homogeneously their temperatures. Concurrently, the structure would be monitored by a camera to deduce the nature of the correlation that exists between changes in the index of refraction and our measurements

Measurement

- **Strain and stress calculation**
  Once we have measured the displacement at several nodes, we can deduce the corresponding strain matrices leading to the calculation of the stresses of the structure. Let's assume that $ux_{i,c}(t)$ denotes the horizontal displacement of floor i at column j while $uy_{i,j}(t)$ denotes its vertical displacement. We define now the strain matrix of floor i and column j at time t as,

$$\varepsilon_{i,j}(n) = \begin{bmatrix} \dfrac{ux_{i+1,j}(t) - ux_{i,j}(t)}{\Delta x_i} & \dfrac{1}{2}\left(\dfrac{uy_{i+1,j}(t) - uy_{i,j}(t)}{\Delta x_i} + \dfrac{ux_{i+1,j}(t) - ux_{i,j}(t)}{\Delta y_i}\right) \\ \dfrac{1}{2}\left(\dfrac{ux_{i+1,j}(t) - ux_{i,j}(t)}{\Delta y_i} + \dfrac{uy_{i+1,j}(t) - uy_{i,j}(t)}{\Delta x_i}\right) & \dfrac{uy_{i+1,j}(t) - uy_{i,j}(t)}{\Delta y_i} \end{bmatrix} \quad 10$$

where,

$$\Delta x_i = < ux_{i+1,j}(t) - ux_{i,j}(t) >_t$$
$$\Delta y_i = < uy_{i+1,j}(t) - uy_{i,j}(t) >_t$$

From the strain matrix described in equation (10) we deduce the stress matrix using the spatial Hooke's law recounted as,

$$\sigma_{i,j}(t) = 2\mu\varepsilon_{i,j}(t) + \lambda tr(\varepsilon_{i,j}(t))I$$

where $\sigma_{i,j}(t)$ is the stress matrix of floor i of column j at time t, μ is the shear modulus, λ is the second Lame parameter, and I the identity matrix. λ and μ can be calculated using the relative Poisson's ratio of the material. The determinant of $\sigma_{i,j}(t)$, $\det(\sigma_{i,j}(t))$, produces the second stress invariant, which has the same value regardless of the coordinate system's orientation.